REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0790

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December, 1994	Final

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1994 Graduate Student Research Program Final Reports, Volume 9, Rome Laboratory

5. FUNDING NUMBERS

6. AUTHORS
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research and Development Labs, Culver City, CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Contract Number: F49620-93-C-0063

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Graduate Student Research Program (USAF- GSRP) is designed to introduce university, college, and technical institute graduate students to Air Force research. This is accomplished by the graduate students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1994

GRADUATE STUDENT RESEARCH PROGRAM FINAL REPORTS

VOLUME 9

ROME LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1994

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

This document is one of a set of 16 volumes describing the 1994 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A & 5B	Wright Laboratory
6	Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center
	Graduate Student Research Program (GSRP) Reports
7	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10	Wright Laboratory
11	Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

1. INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

The numbers of projected summer research participants in each of the three categories are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest; and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1994 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After completion of the 1990 contract, RDL won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

Table 1: SRP Participation, by Year

YEAR	Number of Participants			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years; in one case a laboratory did not fund any additional associates. However, the table shows that, overall, the number of participating associates increased this year because two laboratories funded more associates than they had in previous years.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 44-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous (over 600 annually) individual requesters.

Due to a delay in awarding the new contract, RDL was not able to place advertisements in any of the following publications in which the SRP is normally advertised: *Black Issues in Higher Education, Chemical & Engineering News, IEEE Spectrum* and *Physics Today*.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1994 are shown in this table.

Table 2: 1994 Applicants and Participants

PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	600	192	30
(HBCU/MI)	(90)	(16)	(7)
GSRP	322	117	11
(HBCU/MI)	(11)	(6)	(0)
HSAP	562	133	14
TOTAL	1484	442	55

4. SITE VISITS

During June and July of 1994, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

In previous years, an RDL program representative visited from seven to ten different HBCU/MIs to promote interest in the SRP among the faculty and graduate students. Due to the late contract award date (January 1994) no time was available to visit HBCU/MIs this past year.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, seven HBCU/MI SFRPs declined after they were selected. The following table records HBCU/MI participation in this program.

Table 3: SRP HBCU/MI Participation, by Year

YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6

6. SRP FUNDING SOURCES

Funding sources for the 1994 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1994 SRP selected participants are shown here.

Table 4: 1994 SRP Associate Funding

FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	150	98*1	121*2
USAF Laboratory Funds	37	19	12
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	5	0	0
TOTAL	192	117	133

*1 - 100 were selected, but two canceled too late to be replaced.
*2 - 125 were selected, but four canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

Table 5: 1994 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994
Faculty Members	$690	$718	$740	$740
Graduate Student (Master's Degree)	$425	$442	$455	$455
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391
High School Student (First Year)	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $37/day for graduate students.

Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 58% (178 out of 309) claimed travel reimbursements at an average round-trip cost of $860.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-one percent (78 out of 192) of faculty associates took orientation trips at an average cost of $498. Many faculty associates noted on their evaluation forms that due to the late notice of acceptance into the 1994 SRP (caused by the late award in January 1994 of the contract) there wasn't enough time to attend an orientation visit prior to their tour start date. In 1993, 58 % of SFRP associates took orientation visits at an average cost of $685.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1994 REPORT

The complete set of reports for the 1994 SRP includes this program management report augmented by fifteen volumes of final research reports by the 1994 associates as indicated below:

Table 6: 1994 SRP Final Report Volume Assignments

	VOLUME		
LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, FJSRL, WHMC	6	11	16

AEDC = Arnold Engineering Development Center
FJSRL = Frank J. Seiler Research Laboratory
WHMC = Wilford Hall Medical Center

APPENDIX A – PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP and GSRP associates represent 158 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 46 states plus Washington D.C. and Puerto Rico. Selectees represent 40 states.

GSRP - Applicants came from 46 states and Puerto Rico. Selectees represent 34 states.

HSAP - Applicants came from fifteen states. Selectees represent ten states.

C. Academic Disciplines Represented

The academic disciplines of the combined 192 SFRP associates are as follows:

Electrical Engineering	22.4%
Mechanical Engineering	14.0%
Physics: General, Nuclear & Plasma	12.2%
Chemistry & Chemical Engineering	11.2%
Mathematics & Statistics	8.1%
Psychology	7.0%
Computer Science	6.4%
Aerospace & Aeronautical Engineering	4.8%
Engineering Science	2.7%
Biology & Inorganic Chemistry	2.2%
Physics: Electro-Optics & Photonics	2.2%
Communication	1.6%
Industrial & Civil Engineering	1.6%
Physiology	1.1%
Polymer Science	1.1%
Education	0.5%
Pharmaceutics	0.5%
Veterinary Medicine	0.5%
TOTAL	100%

Table A-1. Total Participants

Number of Participants	
SFRP	192
GSRP	117
HSAP	133
TOTAL	442

Table A-2. Degrees Represented

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	189	0	189
Master's	3	47	50
Bachelor's	0	70	70
TOTAL	192	117	309

Table A-3. SFRP Academic Titles

Academic Titles	
Assistant Professor	74
Associate Professor	63
Professor	44
Instructor	5
Chairman	1
Visiting Professor	1
Visiting Assoc. Prof.	1
Research Associate	3
TOTAL	192

Table A-4. Source of Learning About SRP

SOURCE	SFRP		GSRP	
	Applicants	Selectees	Applicants	Selectees
Applied/participated in prior years	26%	37%	10%	13%
Colleague familiar with SRP	19%	17%	12%	12%
Brochure mailed to institution	32%	18%	19%	12%
Contact with Air Force laboratory	15%	24%	9%	12%
Faculty Advisor (GSRPs Only)	--	--	39%	43%
Other source	8%	4%	11%	8%
TOTAL	100%	100%	100%	100%

Table A-5. Ethnic Background of Applicants and Selectees

	SFRP		GSRP		HSAP	
	Applicants	Selectees	Applicants	Selectees	Applicants	Selectees
American Indian or Native Alaskan	0.2%	0%	1%	0%	0.4%	0%
Asian/Pacific Islander	30%	20%	6%	8%	7%	10%
Black	4%	1.5%	3%	3%	7%	2%
Hispanic	3%	1.9%	4%	4.5%	11%	8%
Caucasian	51%	63%	77%	77%	70%	75%
Preferred not to answer	12%	14%	9%	7%	4%	5%
TOTAL	100%	100%	100%	100%	99%	100%

Table A-6. Percentages of Selectees receiving their 1st, 2nd, or 3rd Choices of Directorate

	1st Choice	2nd Choice	3rd Choice	Other Than Their Choice
SFRP	70%	7%	3%	20%
GSRP	76%	2%	2%	20%

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	116
USAF Laboratory Focal Points	109
USAF Laboratory HSAP Mentors	54

All groups indicate near-unanimous enthusiasm for the SRP experience.

Typical comments from 1994 SRP associates are:

> "[The SRP was an] excellent opportunity to work in state-of-the-art facility with top-notch people."
>
> "[The SRP experience] enabled exposure to interesting scientific application problems; enhancement of knowledge and insight into 'real-world' problems."
>
> "[The SRP] was a great opportunity for resourceful and independent faculty [members] from small colleges to obtain research credentials."
>
> "The laboratory personnel I worked with are tremendous, both personally and scientifically. I cannot emphasize how wonderful they are."
>
> "The one-on-one relationship with my mentor and the hands on research experience improved [my] understanding of physics in addition to improving my library research skills. Very valuable for [both] college and career!"

Typical comments from laboratory focal points and mentors are:

> "This program [AFOSR - SFRP] has been a 'God Send' for us. Ties established with summer faculty have proven invaluable."
>
> "Program was excellent from our perspective. So much was accomplished that new options became viable "
>
> "This program managed to get around most of the red tape and 'BS' associated with most Air Force programs. Good Job!"
>
> "Great program for high school students to be introduced to the research environment. Highly educational for others [at laboratory]."
>
> "This is an excellent program to introduce students to technology and give them a feel for [science/engineering] career fields. I view any return benefit to the government to be 'icing on the cake' and have usually benefitted."

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below (Note: basically the same as in previous years.)

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Laboratory sponsor seminar presentations of work conducted by associates, and/or organized social functions for associates to collectively meet and share SRP experiences.

C. Laboratory focal points collectively suggest more AFOSR allocated associate positions, so that more people may share in the experience.

D. Associates collectively suggest higher stipends for SRP associates.

E. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same arguement was used to raise the faculty and graduate student participation time a few years ago.)

2. 1994 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 109 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	10	30	50	0	20	50	40	0	10	40	60	0	0
AL	44	34	50	6	9	54	34	12	0	56	31	12	0
FJSRL	3	33	33	33	0	67	33	0	0	33	67	0	0
PL	14	28	43	28	0	57	21	21	0	71	28	0	0
RL	3	33	67	0	0	67	0	33	0	100	0	0	0
WHMC	1	0	0	100	0	0	100	0	0	0	100	0	0
WL	46	15	61	24	0	56	30	13	0	76	17	6	0
Total	121	25%	43%	27%	4%	50%	37%	11%	1%	54%	43%	3%	0%

LFP Evaluation Summary. The summarized repsonses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5. a. Benefits of associate's work to laboratory:
    b. Benefits of associate's work to Air Force:
6. a. Enhancement of research qualifications for LFP and staff:
    b. Enhancement of research qualifications for SFRP associate:
    c. Enhancement of research qualifications for GSRP associate:
7. a. Enhancement of knowledge for LFP and staff:
    b. Enhancement of knowledge for SFRP associate:
    c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10. a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
    b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*FJSRL*	*PL*	*RL*	*WHMC*	*WL*
# Evals Recv'd	10	32	3	14	3	1	46
Question #							
2	90 %	62 %	100 %	64 %	100 %	100 %	83 %
2a	**3.5**	**3.5**	4.7	4.4	4.0	4.0	**3.7**
2b	4.0	3.8	4.0	4.3	4.3	4.0	3.9
3	4.2	3.6	4.3	3.8	4.7	4.0	4.0
4	3.8	3.9	4.0	4.2	4.3	NO ENTRY	4.0
5a	4.1	4.4	4.7	4.9	4.3	3.0	4.6
5b	4.0	4.2	4.7	4.7	4.3	3.0	4.5
6a	**3.6**	4.1	**3.7**	4.5	4.3	**3.0**	4.1
6b	3.6	4.0	4.0	4.4	4.7	3.0	4.2
6c	3.3	4.2	4.0	4.5	4.5	3.0	4.2
7a	3.9	4.3	4.0	4.6	4.0	3.0	4.2
7b	4.1	4.3	4.3	4.6	4.7	3.0	4.3
7c	3.3	4.1	4.5	4.5	4.5	5.0	4.3
8	4.2	4.3	5.0	4.9	4.3	5.0	4.7
9	3.8	4.1	4.7	5.0	4.7	5.0	4.6
10a	4.6	4.5	5.0	4.9	4.7	5.0	4.7
10b	4.3	4.2	5.0	4.3	5.0	5.0	4.5
11	4.1	4.5	4.3	4.9	4.7	4.0	4.4
12	4.1	3.9	4.0	4.4	4.7	3.0	4.1
13a	**3.8**	**2.9**	4.0	4.0	4.7	**3.0**	**3.6**
13b	**3.8**	**2.9**	4.0	4.3	4.7	**3.0**	**3.8**
14	4.5	4.4	5.0	4.9	4.7	4.0	4.5

3. 1994 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 275 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

1. The match between the laboratories research and your field:		4.6
2. Your working relationship with your LFP:		4.8
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.5
5. Lab readiness for you: LFP, task, plan:		4.3
6. Lab readiness for you: equipment, supplies, facilities:		4.1
7. Lab resources:		4.3
8. Lab research and administrative support:		4.5
9. Adequacy of brochure and associate handbook:		4.3
10. RDL communications with you:		4.3
11. Overall payment procedures:		**3.8**
12. Overall assessment of the SRP:		4.7
13. a. Would you apply again?	Yes:	85%
b. Will you continue this or related research?	Yes:	95%
14. Was length of your tour satisfactory?	Yes:	86%
15. Percentage of associates who engaged in:		
a. Seminar presentation:		52%
b. Technical meetings:		32%
c. Social functions:		03%
d. Other		01%

16. Percentage of associates who experienced difficulties in:

a. Finding housing:	12%
b. Check Cashing:	03%

17. Where did you stay during your SRP tour?

a. At Home:	20%
b. With Friend:	06%
c. On Local Economy:	47%
d. Base Quarters:	10%

THIS SECTION FACULTY ONLY:

18. Were graduate students working with you?	Yes:	23%
19. Would you bring graduate students next year?	Yes:	56%

20. Value of orientation visit:

Essential:	29%
Convenient:	20%
Not Worth Cost:	01%
Not Used:	34%

THIS SECTION GRADUATE STUDENTS ONLY:

21. Who did you work with:

University Professor:	18%
Laboratory Scientist:	54%

4. 1994 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

The summarized results listed below are from the 54 mentor evaluations received.

1. Mentor apprentice preferences:

Table B-3. Air Force Mentor Responses

		How Many Apprentices Would You Prefer To Get ?			
		HSAP Apprentices Preferred			
Laboratory	***# Evals Recv'd***	***0***	**1**	**2**	**3+**
AEDC	6	0	100	0	0
AL	17	29	47	6	18
PL	9	22	78	0	0
RL	4	25	75	0	0
WL	18	22	55	17	6
Total	**54**	**20%**	**71%**	**5%**	**5%**

Mentors were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

2. Mentors involved in SRP apprentice application evaluation process:
   a. Time available for evaluation of applications:
   b. Adequacy of applications for selection process:
3. Laboratory's preparation for apprentice:
4. Mentor's preparation for apprentice:
5. Length of research tour:
6. Benefits of apprentice's work to U.S. Air force:
7. Enhancement of academic qualifications for apprentice:
8. Enhancement of research skills for apprentice:
9. Value of U.S. Air Force/high school links:
10. Mentor's working relationship with apprentice:
11. Expenditure of mentor's time worthwhile:
12. Quality of program literature for apprentice:
13. a. Quality of RDL's communications with mentors:
    b. Quality of RDL's communication with apprentices:
14. Overall assessment of SRP:

	AEDC	*AL*	*PL*	*RL*	*WL*
# Evals Recv'd	**6**	**17**	**9**	**4**	**18**
Question #					
2	100 %	76 %	56 %	75 %	61 %
2a	4.2	4.0	**3.1**	**3.7**	**3.5**
2b	4.0	4.5	4.0	4.0	3.8
3	4.3	**3.8**	**3.9**	**3.8**	**3.8**
4	4.5	**3.7**	**3.4**	4.2	**3.9**
5	**3.5**	4.1	**3.1**	**3.7**	**3.6**
6	4.3	3.9	4.0	4.0	4.2
7	4.0	4.4	4.3	4.2	3.9
8	4.7	4.4	4.4	4.2	4.0
9	4.7	4.2	3.7	4.5	4.0
10	4.7	4.5	4.4	4.5	4.2
11	4.8	4.3	4.0	4.5	4.1
12	4.2	4.1	4.1	4.8	3.4
13a	**3.5**	**3.9**	**3.7**	4.0	**3.1**
13b	4.0	4.1	3.4	4.0	3.5
14	4.3	4.5	3.8	4.5	4.1

5. 1994 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 116 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

1. Match of lab research to you interest:		3.9
2. Apprentices working relationship with their mentor and other lab scientists:		4.6
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.1
5. Lab readiness for you: mentor, task, work plan		3.7
6. Lab readiness for you: equipment supplies facilities		4.3
7. Lab resources: availability		4.3
8. Lab research and administrative support:		4.4
9. Adequacy of RDL's apprentice handbook and administrative materials:		4.0
10. Responsiveness of RDL's communications:		3.5
11. Overall payment procedures:		3.3
12. Overall assessment of SRP value to you:		4.5
13. Would you apply again next year?	Yes:	88%
14. Was length of SRP tour satisfactory?	Yes:	78%
15. Percentages of apprentices who engaged in:		
a. Seminar presentation:		48%
b. Technical meetings:		23%
c. Social functions:		18%

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Ph.D.
Professor

Slawomir J. Marcinkowski
Ph.D. Student

John Carlo Bertot
Ph.D. Student

School of Information Studies
4-293 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100

Final Report for:
Graduate Summer research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Slawomir J. Marcinkowski and John Carlo Bertot
School of Information Studies
Syracuse University

Abstract

The current global economic climate is such that a nation acquires and maintains its wealth, prosperity, and strength predominantly through trade. It is necessary, but no longer sufficient, for a nation to possess a strong military function in a global marketplace. This new emphasis on national competitiveness in trade places Rome Laboratory, as well as other federal laboratories, at an important crossroads. On the one hand, a military advantage requires continual technological superiority. On the other hand, Rome Laboratory needs to facilitate national economic development through the transfer of its technologies into the marketplace. These developments serve to highlight the importance of a proactive technology transfer process within Rome Laboratory. However, a proactive operation is difficult to put into action without forthcoming budgetary and personpower increases. This study focuses on a low-cost alternative: to use MOSAIC on the Internet and the World Wide Web to promote the transfer and commercialization of technology. An electronic system was developed allowing access on various technology transfer information and databases to the private sector, as well as Rome Laboratory and other Air Force and public sector users. This report describes these efforts, underlying reconceptualizations, design and implications of the electronic system.

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Slawomir J. Marcinkowski and John Carlo Bertot

Introduction

The current global economic climate is such that a nation acquires and maintains its wealth, prosperity, and strength predominantly through trade. It is necessary, but no longer sufficient, for a nation to possess a strong military function in a global marketplace. This new emphasis on national competitiveness in trade places Rome Laboratory (RL), as well as other federal laboratories, at an important crossroads. On the one hand, a military advantage requires continual technological superiority. On the other hand, RL needs to facilitate national economic development through the transfer of its technologies into the marketplace. These developments serve to highlight the importance of a proactive technology transfer operation within RL. However, a proactive operation is difficult to put into action without forthcoming budgetary and personpower increases. This study attempts to identify a low-cost alternative: to use the Internet to promote the transfer of technology.

RL is currently positioning itself to meet the challenges precipitated by the emphasis on the technology transfer process. This process, if viewed as a life cycle, minimally contains the following components:

Technology research

Technology development

Technology announcement

Technology marketing

Fulfilling technology information requests

Making the technology available for outside use

Technology utilization, and

Technology commercialization.

In addition, technology transfer occurs within political and social climates that may impose certain constraints. The overriding condition in all of these considerations is, of course, first of all the Air Force

mission. In that sense technology transition within the Air Force is of primary importance, followed by technology transfer to the private sector, hopefully resulting in technology commercialization. The latter term incorporates such desirable events as successful technology transfer, technology utilization, profitability of the technology when marketed by a firm, hopefully resulting in the creation of jobs, adding value, contributing to the regional and national economies, and adding overall to the Gross Domestic Product.

Many of these technology transfer tasks are mandates and specified by legislation such as the Stephenson-Wydler Act, the Technology Transfer Act, and others. There are also certain limitations and constraints on the technology transfer process such as export control regulations intended to help only U. S.-owned companies.

Technology transfer deliverables could be classified as: products, knowledge, advice, know-how, and facilities use. These deliverables are not unique to RL, but can be envisioned to apply to all federal laboratories. They can be broken down into a set of specific and clearly definable technology transfer activities may include the following:

Cooperative Research and Development Agreements (CRDAs)

Educational Partnerships

Exhibits

Conferences

Presentations (by the Technology Transfer Office or scientists and engineers)

White papers

Information exchange

Inquiries (face-to-face, by phone, written)

Publications

Advertising

Public relations

Communication with company representatives

Electronic dissemination

Others.

The present focus for this study is on just one of the above activities, electronic dissemination of information about technology.

The authors reviewed the technology transfer process at RL and revisited this process by reconceptualizing the underlying assumptions pertaining to basic principles of marketing, diffusion, information retrieval and networking. The key driver of this reconceptualization was that of a user or target audience orientation. From this perspective a model was developed that provides the desired electronic linkage between RL and targeted private sector firms.

Reconceptualization

Four different approaches to a reconceptualization of the technology transfer process, especially when this process occurs electronically, were advanced. Each is described below:

A. Marketing. All marketing efforts are based on the basic premise that there is a specific audience of consumers. This is a set of specific firms within an industry or the private sector, which have needs that can be filled by firms operating within a specific market. In the this case, RL constitutes such a firm that can fill the demand for specific technologies that can be marketed and commercialized by private sector firms.

One can identify three main policies of orientation within the marketing effort: customer orientation, product orientation and profit orientation. The latter orientation is not an appropriate orientation for RL. A product orientation is predicated on the view that consumers will recognize and appreciate products for superior merit and bestow their patronage on firms--in this case the product is research and development. In the present case though, it must be realized that, in general, this product is not designed by RL having private sector firms in mind foremost. Rather, the key customer is the Air Force and the product and the product orientation are determined by the Air Force mission. An exclusive product orientation effort is therefore not wise for RL, but the product certainly determines a potential market that can be tapped within the private sector. It is therefore apparent that an appropriate focus for RL in terms of marketing

orientation is a customer orientation.

In this setting, a customer orientation denotes (a) an attitude and a pattern of conduct, as well as (b) the extent to which RL tries to determine what its customers want and then gives them what they want. Granted, it is often difficult and complex to discover what the customer wants. For example, in such efforts one recognizes that customer preferences vary widely and that customers want something only once they are given concrete choices, i. e. they are incapable of conceiving possible new products. Potential customers often also request features that are incompatible with product capabilities.

The basic challenge faced by RL is to identify those needs and provide a linkage between RL and the customer, i. e. a firm interested in a specific RL technology. To maximize this linkage and relationship to customers, RL needs to form hypotheses and understanding about these present and future potential customers. Included should be questions such as:

What kinds of things affect customer behavior?

Through which channels (advertising, face-to-face contacts, publications, etc.) can customers be reached by RL?

What is the degree or strength of need or desire for the product?

What are the appropriate appeals (or arguments) to which customers are most responsive?

What is the customer's responsiveness to different types of sales devices, i. e. their ability to be influenced by technology transfer discussions by the Technology Transfer staff, engineers and scientists?

After these questions have been answered, the marketing dimension entails five general activities:

1. Identifying and selecting the type of customer that RL chooses to cultivate and learning that set of firms' needs and desires.

2. Designing products, know-how and services, to include facilities, that RL can transfer in conformity with customer desires.

3. Persuading customers to acquire and adopt RL's products, know-how and services.

4. Displaying, moving and to some extent storing products, know-how and services after they have

been developed at RL.

5. Identifying dual use technologies and applications.

In designing its products, acquiring and developing know-how, and deciding how much and what specific services to offer, RL will unquestionably benefit from having a clear picture of its *target customer*. The present project attempts to provide a linkage, an *electronic marketing channel*, between these target customers and RL.

B. Diffusion. Diffusion is another conceptualization that guided the study. Diffusion is the social process by which an innovation is communicated through certain channels over time among members of a social system (Rogers, 1983). This process clearly resembles the task at hand for RL in the technology transfer process: i. e. a technology or an innovation within RL, needs to be communicated to a set of firms (members of a social system) within a particular industry. The communication channel chosen for the present study is that of electronic messaging via the Internet. The authors view such electronic diffusion means as timely, cost-effective, quick and targeting specific means to reach the target customer described above.

Many other dimensions could be addressed here that determine the speed with which such diffusion occurs. Obviously the quicker the target audience can be reached and the significance of the technology to be transferred can be communicated, the more effective the diffusion process. The rate of diffusion or adoption of that technology is, however, determined by the characteristics of the innovation (e. g., the relative advantage, compatibility, complexity, trialability and observability).

The communication channel chosen, electronic messaging via the Internet, determines in part the success of making a successful linkage with the target customer. General diffusion principles posit that, ideally, successful communication occurs in a face-to-face setting with the target audience or customer. This, however, is often impracticable or too expensive. Substitutes to face-to-face settings are chosen and the mass media come into play. The use of mass media in such efforts, as in advertising, is often a vague undertaking, as one cannot be sure that one's message truly reaches the intended target. Often a compromise between the two extremes is desirable, e. g., the use of mass media, followed up by face-to-

face meetings with smaller groups, etc. Electronic messaging via the Internet comes close to this ideal, i. e. using a cost-effective mass medium and still reaching specific individuals. Moreover, the Internet's use makes it possible to be interactive with specific individuals within the target group and thus allowing *almost* face-to-face interaction. This potential for interactivity certainly makes the medium highly attractive as requests, needs, etc. certainly can be customized. Moreover, such interactivity makes possible the often missing *feedback* in the technology transfer process. Such feedback allows us to shape and additionally customize the very next step in the diffusion and communication process. Such customization is almost impossible when viewing the diffusion process via advertising as a communication channel.

C. Information Retrieval. In an electronic system, such as using the Internet, it is obvious that the users can readily retrieve information once the information is stored on the system. There are many data bases available at RL that could be accessed through the Internet, and many of those pertain to technology transfer and related issues. The authors made every effort to identify such data bases and make them available in a format conducive to outside users, hopefully firms that might acquire a particular product, know-how or service. The design of such information, data bases, etc. often determines their successful use. In this sense, a customer or *end-user* orientation was applied in the design of the system. Unless such design is inviting, encouraging, timely, informative and user-friendly, the success of the system is highly questionable (Taylor, 1986).

D. Networking. The topic of networking as a conceptualization embeds all three of the above topics, i. e. marketing, diffusion and information retrieval. Without these three, networking would be impossible. Networking in this sense goes beyond the traditional means of reaching a target customer. George Kozmetsky, director of the IC2 Institute, Austin, Texas, stated at a recent conference on commercialization of technology from federal laboratories that, "The secret to successful commercialization of technology developed in the federal labs is networking, networking, networking (Kozmetsky, 1993)." This importance was already demonstrated in an empirical NSF-funded study by Wigand focusing on technology transfer issues within the microelectronic industry and the role of industry, government and universities (Wigand and Frankwick, 1989).

Networking denotes interaction, feedback, customizing and creating a dialogue to a degree that otherwise

would not exist. In that sense and in this particular case, RL may conceive of transferring technology, at least in the beginning stages, as a partially cooperative means with potential target customers within industry. Networking would suggest the creation of Listservs, electronic bulletin boards (discussion groups), direct electronic inquiries that might be shared with the rest of likewise interested individuals/firms around a particular product, know-how or technology. Such individuals within firms appear to be quite interested in such dialogue as a technology emerges, although it is understandable that such dialogue would be reduced or cease entirely once the technology becomes highly marketable or commercializable due to competitive reasons. But even if RL could speed up the development and diffusion of a technology to such a point and if interaction would then discontinue, RL still would have provided a most valuable service to the development of that technology and industry. Typically, such networks do not come about by themselves. Such interaction must be fostered, encouraged , managed and nurtured by knowledgeable professionals within RL, perhaps among the Technology Transfer staff or the Directorate's Technology Transfer Focal Points. Networking, in this sense, is conceived as the highest level of technology transfer along the electronic medium of using the Internet. Lastly, networking as envisioned here, is characterized as rating high in terms of customer interaction, as well as high in terms of information content, implying that RL has the means to get closer to the customer. Moreover, it is interactive, multifaceted, customized, and on demand with time delay and time zone shifts playing merely a minor role.

Methodology

Various methods were employed to investigate and research the task at hand. The authors conducted interviews with numerous individuals within RL and with non-RL employees. Interviews were conducted with the members of the RL Technology Transfer Office staff, including RL elements at Hanscomb AFB, as well as several engineers and scientists. One engineer was interviewed in greater depth, as his ACT technology was highlighted for demonstration purposes on the Internet, i. e. the World Wide Web (WWW) at RL. This particular technology was demonstrated to show the various capabilities and cross-linkages possible using MOSAIC as the software to navigate on the WWW. In addition, several branch chiefs were interviewed as well as several area company representatives (Kaman, TRW and Booze, Allen & Hamilton).

Moreover, in-depth interviews were conducted with the officers of the Photonics Development Corporation in Rome, NY, as well as members of the RL Library. The authors interviewed appropriate, selected private sector firms in terms of their willingness, availability and readiness to enter the Internet and WWW world. The effects of the interviews were surprising to some extent as we saw a *snowball effect* emerging. That is the discussion of our interests in the electronic transfer process generated numerous additional suggestions, existing and potential services, as well as data bases. All in all, a surprising synergy was created between the research team and RL members engaged in research and technology transfer.

The Internet and the World Wide Web

The Internet is a very fast growing organization over the last several years. No one knows for sure the exact number of users, as the Internet itself is an organization of loosely coupled networks. An August 4, 1994 report of the Internet Society, the organizing body representing the Internet, reports dramatic growth in 1994. The latest Internet Society measurements reveal that there are worldwide 3.2 million reachable machines. This is an increase of 81 percent for the past year and represents an even steeper than normal increase over the past six months. Indeed, one million new hosts were added during the first six months of 1994. Of the world-wide total, the U. S. connections account for 63 percent (Internet Society, 1994). The fastest currently growing segment on the Internet are businesses, again demonstrating the potential to reach the private sector via the Internet. Although it is difficult to specify a threshold at which a critical mass of business Internet users has been reached, it is possible today to reach most major firms via the Internet. Small and medium sized firms are representative of the vast majority of firms signing on the Internet (Cf., e. g., with "Going *I-way?*," 1994; McBride, 1994).

An important caveat needs to be expressed here: Just getting a presence on the Internet is not enough. Just putting one's server on the Internet is analogous to hanging a shingle on a rural road. If one is selling something and has something important to offer that is intended to reach a large audience, it is important to have not just a shingle, but a major advertising board on the *main drag*, i. e. a major highway.

The Internet's rapid proliferation suggests that organizations in the public and private sectors want to take advantage of the Internet's interactive or two-way capabilities. In turn, this suggests the establishment and

exploitation of visible and invisible social and physical networks linking and bridging public and private sector organizations.

The appropriate use of the Internet may result in such desirable outcomes as: obtaining support from vendors, participating in joint development, collaborative research projects, providing customer support, marketing and product/service/know-how distribution. Due to space limitations it is impossible here to highlight various features of a value chain derived from appropriate Internet use. This value chain can be segmented into internal operations, inputs from suppliers, and customer relations. The Internet in conjunction with other appropriate technology and software, such as the World Wide Web and MOSAIC, lends itself ideally to provide an effective, quick and responsive interactive linkage between RL and its customers, i. e. various private sector firms (Cf., e. g., with Blattberg et al., 1994; Dallaire, 1993). Having a home page on the World Wide Web is not sufficient; the information content must reflect the potential user's information need. The authors have incorporated into the technology transfer home page the information that is most sought after by potential adopters of RL's technologies. The nature of this linkage is ideal in terms of cost-effectiveness and time-based considerations. Moreover, it provides the potential for feedback, the essential component in any effective communication effort.

Data Bases and Electronic Features Available Via the Internet and WWW

During the summer months of 1994 the authors designed, built and applied various data bases and services of interest in the technology transfer process. As mentioned initially, our efforts were driven by an overall customer/user orientation. The designed system uses MOSAIC on the WWW, is fully accessible to outside, i. e. non-RL, users, as well as RL users and features the following:

What's Happening in Tech Transfer at Rome Lab?

Technology Transfer at Rome Laboratory

Hot Products

Technology Transfer Inquiries

A Model CRDA

CRDA Data Base

Educational Partnerships

Fact Sheets of Directorates

Facilities

Skills at Rome Laboratory Database

Rome Laboratory Staff Directory

The Reliability Engineer's Tool Kit

Patents Granted at Rome Laboratory

The management, use and ownership of these databases poses an intriguing question. It is very important that data bases are kept up-to- date and various current features such as *Hot Products* and *What's Happening in Tech Transfer at Rome Laboratory* feature products and information that truly is current. Otherwise, the system designed will be of little use to the outside, and RL users will recognize quickly that it is not a timely device to find out information about RL and its technologies.

The authors suggest that the ownership of this system resides with the Rome Laboratory Technology Transfer Office. Access to various features should be categorized such that it is available to all users (including the private sector) or only the Rome Laboratory Technology Transfer staff or only the RL Technology Transfer staff and all RL scientists and engineers. We recommend that the gatekeepers of information as well as the implementors of information into the system are the various Technology Transfer Focal Points within each directorate.

Implications

The system designed clearly demonstrates that the technology transfer capabilities can be achieved via electronic means. This elegant design derives value for Rome Laboratory through the timely and appropriate alignment of information technology already in place (personal computers, high-speed networks, software such as MOSAIC, and access to the Internet and WWW), Rome Laboratory's goals and strategies with regard to technology transfer and its larger role within the Armed Forces and society, as well as the appropriate alignment with existing *business processes*, especially with regard to technology transfer. Through the appropriate alignment and organizational fit, it is possible to generate added value

for Rome Laboratory. It is of considerable importance, though, to realize that this designed process occurring electronically cannot be arrested by peripheral traditional processes that might be paper-based, and still assume that the value chain continues delivering the same value. Appropriate electronic alignment must occur in such peripheral processes as well.

One must realize also that implementing this effort, even though highly cost-effective when compared to more traditional means, e.g., advertising, does not come cheaply. The system as conceived must be taken on by an individual or a team of individuals knowledgeable of the technology transfer process, RL, private sector operating practices, and probably some technology background and competence would be helpful as well. The task at hand is not just the mounting of the system on the WWW, but the effort must be nurtured, customized and *mothered*. Especially, the involvement of key individuals in private sector firms with an interest in technology transfer in such electronic organizational forms as Bulletin Boards, Listservs, etc. requires sensitivity, understanding and a sense of managing groups from a group dynamics perspective. This is time-consuming and requires patience in real life, but especially when done electronically. Numerous examples, however, abound demonstrating success in such efforts.

References

Blattberg, Robert C., Glazer, Rashi, and Little, John D. C. (Eds.). *The Marketing Information Revolution*. Boston, MA: Harvard Business School Press, 1994.

Dallaire, Rene M. Data-Based Marketing for Competitive Advantage. *Data Resource Management*, Spring, 1993, pp. 46-50.

Going *I-way?* [Cover Story]. *PC Computing*, September, 1994, pp. 120-169.

Internet Society Press Release, August 4, 1994.

Kozmetsky, George. "Commercialization of Technology" (Keynote Address). Presented at the Conference on Commercialization of Technology within Federal Laboratories, Santa Fe, New Mexico, March, 1993.

McBride, James. Marketing MOSAIC: The War Has Begun. *Internet World*, October, 1994, pp. 40-43.

Rogers, Everett M. *Diffusion of Innovations* (Third Edition). New York: Free Press, 1983.

Taylor, Robert S. *Value-Added Processes in Information Systems*. Norwood, NJ: Ablex, 1986.

Wigand, Rolf T. and Frankwick, Gary L. Inter-Organizational Communication and Technology Transfer: Industry-Government-University Linkages. *International Journal of Technology Management*, 1989, 4(1), pp. 63-76.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas
Graduate Student
Department of Electrical and Computer Engineering

University of Arizona
1230 E. Speedway Blvd.
Tucson, AZ 85721

Final Report for:
Graduate Summer Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air force Base, DC

and

Rome Laboratory

August 1994

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas
Graduate Student
Department of Electrical and Computer Engineering
University of Arizona

Abstract

In investigating front end model development, an environment is described that allows for model insertion and evaluation. The preformal stages of the model will be represented by an English language verb phrase. This representation is sufficiently detailed to serve as the basis for model construction and yet sufficiently "soft" to support knowledge acquisition during model construction. This paper establishes the adequacy of this representation.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

Introduction

There are problems in the current modeling process associated with model development. Although the "back-end" of the model-engineering process (Fishwick 1990) is well-supported by software tools, the same can hardly be said for the "front-end" -- model creation, construction, or repository-based synthesis.

This paper suggests that such "front-end" problems can be substantially ameliorated by adopting extraction and aggregation techniques that employ decomposition to identify essential components for model construction. The particular objectives for model construction that we wish to address are extracted through an English Language Interface, hereafter referred to as ELI.

The goal of this approach is to reduce ambiguity between verbal descriptions of requirements and identifying essential model construction components. Requirements here are represented by verb phrases. Each one has the potential to become a complete discrete event model. This natural language process describes a method of constructing models by users with a limited or nonexistent formal modeling or programming background.

Overview

The ELI allows model specification in terms of a verb phrase. It consists of a verb, noun, and modifier. An example might be "build car quickly." In this case the verb is build, the noun is car, and the modifier is quickly. The verb "build" would be parameterized by the noun "car." The noun specifies the domain that the verb operates or acts on. The term "parameterized by the noun" is used in our context to mean the descriptive components (parameters) of the noun that provide further detail about it's domain. For example, in "build car quickly," the car would have parameters such as length, weight, etc. The modifier "quickly" adds range to the parameterization. This is shown in Figure 1.

A Proposed Verb Phrase Decomposition System Concept

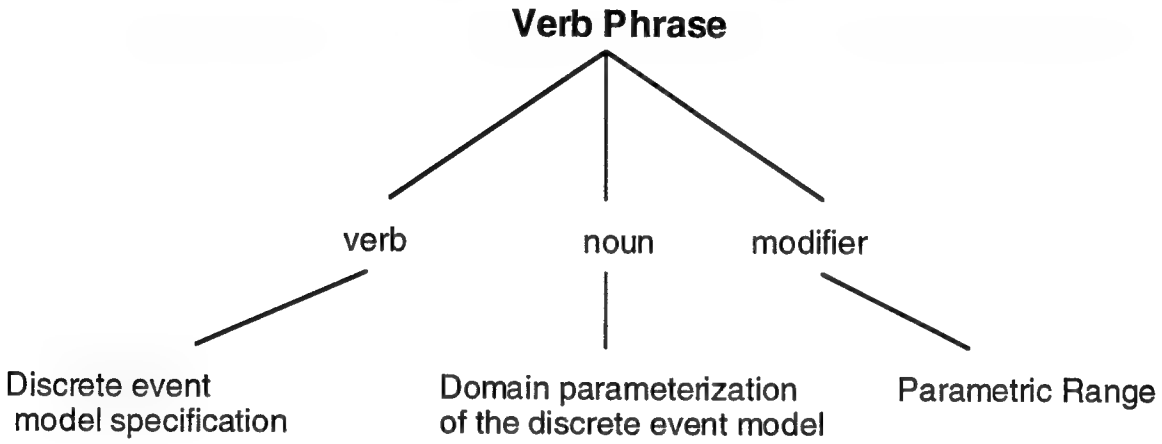


Figure 1. Verb Phrase Decomposition

Conceptual realization of a model from a verb phrase ties in closely with Checkland's (1990) idea of having a verb express the root definition, or core purpose, of a system. He explains "That core purpose is always expressed as a transformation process in which some entity, the 'input', is changed, or transformed, into some new form of that same entity, the 'output'." Using the verb as a discrete event model template is an extrapolation of this idea.

With the basic discrete event model defined from the root definition, or verb, the noun of the verb phrase gives the user a working domain. The verb phrase's noun defines the domain of possible action. For example, "build car" parameterizes the domain in units of the mean time to build a car. This might be hours for a car. When applied to a house - "build house," this might be months.

The modifier adds focus to the noun's domain. For example, in "build car quickly," the modifier "quickly" optimizes all car building parameters. The effect of the modifier could be quantified at the user's discretion.

With each model represented by a verb phrase, a system could be decomposed into a semantic representation as shown in Figure 2. In this case, a hierarchical car production system goes from production planning to practical assembly line issues - with each step represented by a verb phrase. The focus here is not so much on the system or its structure, but on how the ELI allows semantic representation of the system.

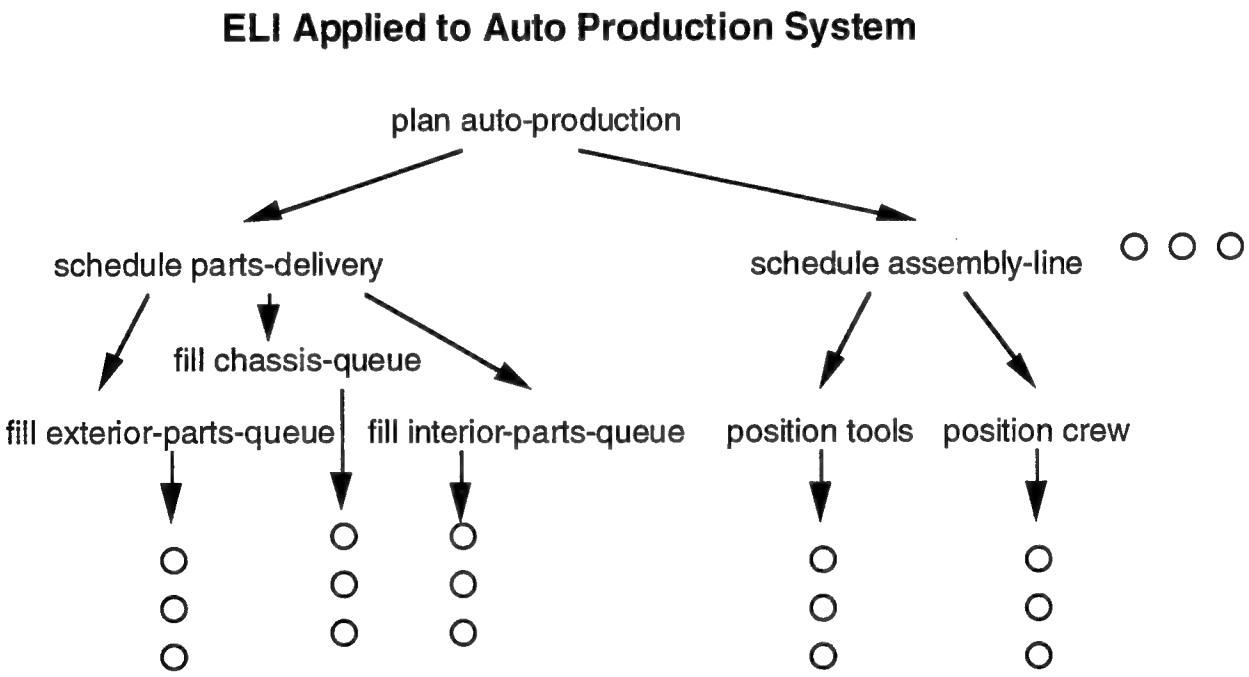


Figure 2. ELI representation of car production system hierarchy

The focus here is on description of the front end model construction process. Each system function would be modeled as a verb phrase. The verb represents the basic action to be taken, and forms the basis of the discrete event model. A noun then defines the domain. The modifier defines how the action will be performed within the domain. The verb-noun-modifier triplet defines the verb phrase. Now we will look at that definition in more detail.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

Description of Verb Phrase Extraction Process

The following is a complete description of how the verb phrase is constructed. Examples of verb definition, noun parameterization, and modifier parameters of limitation are given.

Verb Description

Using Fernald's (1957) definition, the verb is transitive in requiring the verb phrase's noun to complete its meaning. It is principal in expressing the act to be done. And, its voice is either active or passive because the subject can either be acting or acted upon. Using Fernald's definition, we can develop a description of verb use in this system.

Verb Definition

Verbs are used to describe the two main categories of system behavior, which we propose to be production and consumption. These verbs form the basis for each model. We also propose the verb to represent the basic discrete event model template.

The first class of verbs, which we call producers, describe value added activity. This could be anything from growing crops to stamping sheet metal into automobile body panels. A producer is active. It models something that acts on the noun of the verb phrase.

Consumption is less straightforward. Consumption can be sensory or physical. Sensory consumption is defined as what we consume or observe. An example of sensory consumption could be the use of the five senses in observing one's environment. Physical consumption might be taking in nutrients for processing into energy.

Verb definition is performed by descending a verb hierarchy until a satisfactory representation is found. For example, if one's goal is building, we descend the verb's hierarchy until choosing "build" under the producer class. This is shown in Figure 3.

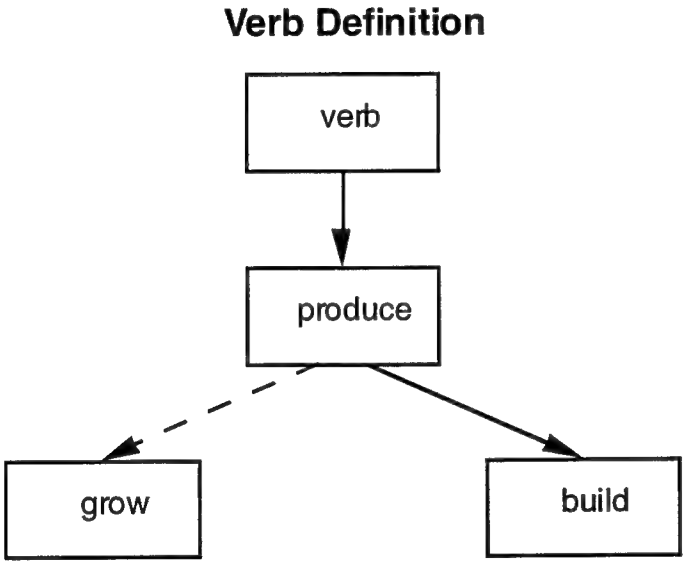


Figure 3. Producer verb build is chosen

Verbs, regardless of their class, are tied to reality by their domain. For example, the verb "build" can be used to build a car with its variables describing the action. In this case of build, the variables might be cycle time, feed rate, idle time, etc. The user controls the modification of these variables.

The variables of a verb might be in the form of an array. For example, the verb "build" might have the following variables:

build = s[feed-rate][cycle-time][idle-time]

In this case, the variables of "build" are:

- feed-rate
- cycle-time
- idle-time

And the numerical parameter ranges would be filled in when the noun, and its accompanying domain, is chosen.

We are proposing that each verb is represented by a discrete event model shell with state variables that can be used to describe its behavior. These variables would be open for user modification. In order to simplify the understanding of verb use in this system, we also propose that they fit into a taxonomy.

Verb Taxonomy

This breakdown of verb classes gives us structure in thinking about verbs and the actions that they represent. As shown in Figure 4, the actions of producers and consumers are different. We believe this breakdown provides the user added structure when describing system requirements for model construction.

Jerry M. Couretas

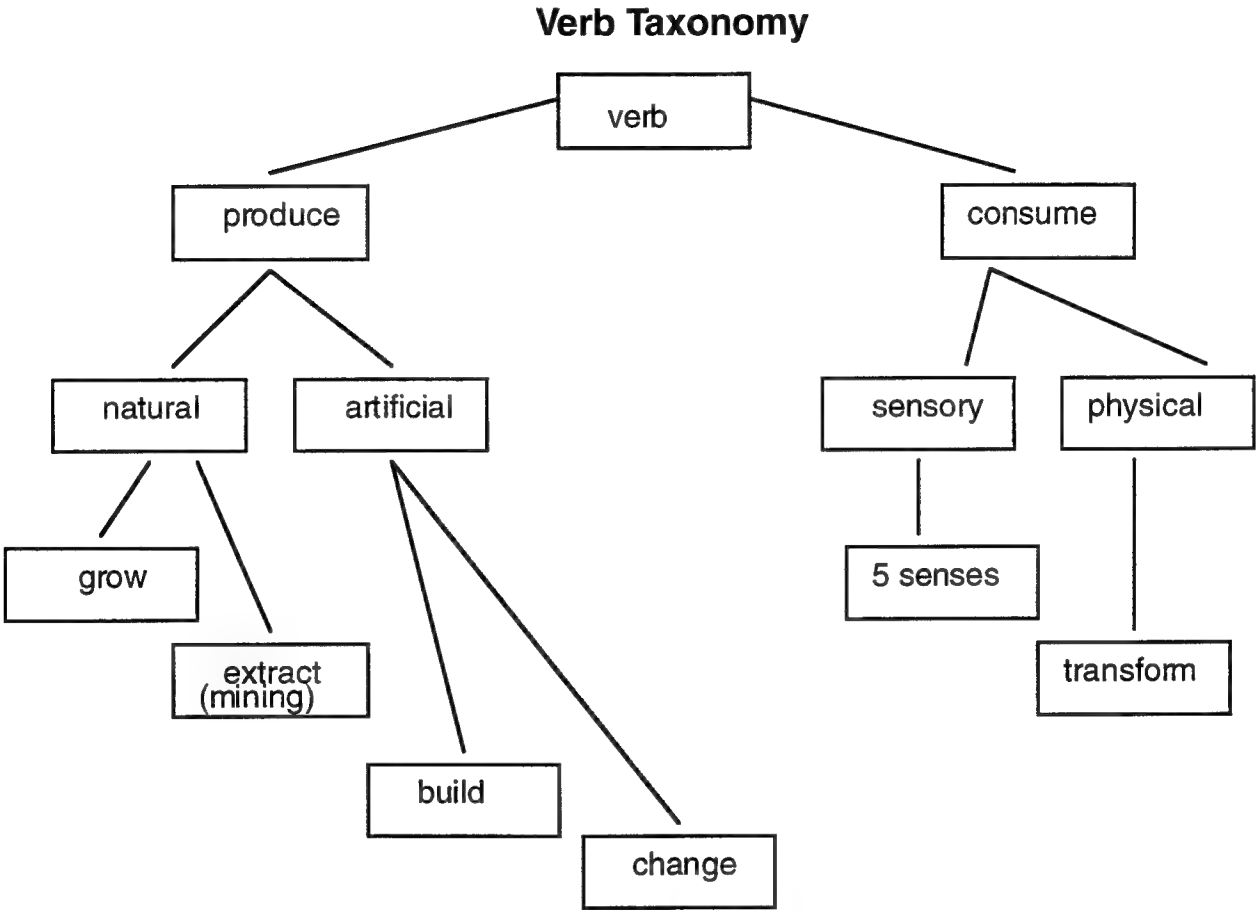


Figure 4. Taxonomic breakdown of verb into classes of production and consumption

Verb Classification

The key to verb classification is what transforms it and how. Some potential verb representing transformations might be:

State Change Transformation of Producers and Consumers

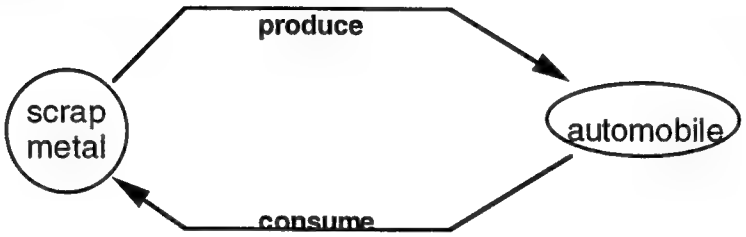


Figure 5. Verb Classification State Changes

In looking at the above diagram, we see the following verb transformations:

Producers

Value is added to the initial state of the noun. The end state is a transformation from scrap metal to a useful automobile.

Consumers

The final state of the noun is transformed from its initial state of an automobile to scrap metal.

Producers and consumers could work together within a system. This classification system gives the user a simple way of classifying almost any process in a system. In the car production example of Figure 2, an automotive assembly line consumes component parts and produces complete automobiles. We know that we are dealing with a relatively high level system when it has both classes of verb in its definition. Decomposing this system, as shown in Figure 6, shows how the different components act as consumers and producers.

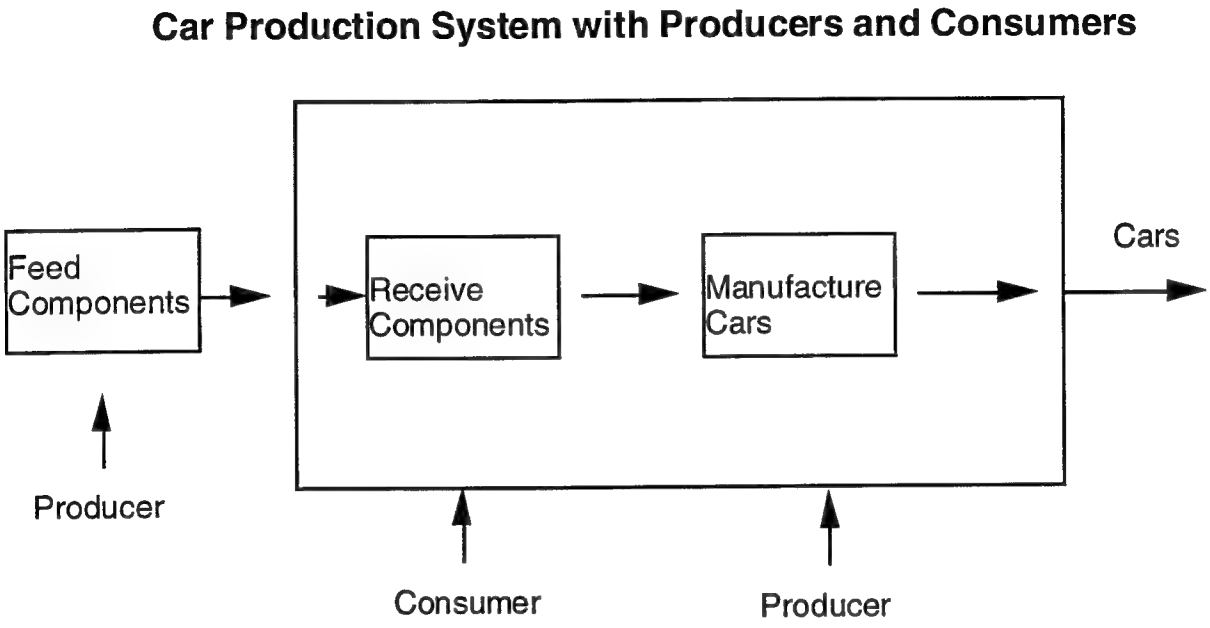


Figure 6. Example system with different model types

Looking at the above verb classification, we see that the main difference between producers and consumers is their relationship to the noun. A producer acts on the noun and a consumer is acted upon by the noun.

Produce
material -> finished product
location(1) -> location (2)
information -> report
layman -> professional
undeveloped land -> building

Consume
sensory data -> information
material -> energy
land -> contaminated soil
young -> old (physical capability)
current -> obsolete (practicality)

Suh(1990) has a comparable method where each functional requirement maps into a set of data parameters. Functional requirements account for the functional space of the design, while the data parameters account for the physical space. The mapping between these two is determined by the design, or verb phrase system representation. The ELI is similar in that each verb represents a functional requirement. And, the verb's domain is defined by data parameters. Verb name specification would result in the system accessing a database of verbs along with their parameterization as shown below:

verb = < param1, param2, ... >

build = < feed-rate, cycle-time, queue-time, ... >

In this case, the verb "build" is parameterized by feed-rate, cycle-time, queue-time, etc. This would be the knowledge based component of the system in that whenever a verb is called, it would be parameterized as previously defined. The user would then be free to modify the parameters given.

The verb represents one of the system's actions. It also represents a basic discrete event model with an empty set of descriptive variables. These are open for user modification, and would be parameterized by the noun.

Noun Description

The noun is the second component for construction of the verb phrase. The noun serves as the object of the verb phrase; and thereby establishes its domain parameterization. The noun's definition and taxonomy are presented next.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

Noun Definition

The noun defines the domain in which the verb works. Each noun has a different domain representation, and thus has a different parameterization. The noun's domain knowledge should provide the parameterization necessary to complete the model.

The noun quantifies the verb's action parameters. For example, the verb "build" could be parameterized as follows when applied to the noun "car":

build = s[feed-rate][cycle-time][idle-time]
build = s[5][3][1]

This means that the verb phrase "build car" breaks down to the following numerical parameterization:

feed-rate	=	5
cycle-time	=	3
idle-time	=	1

The noun's domain will define the numerical parameters of the verb's model. Another view of the noun is presented by its taxonomy.

Noun Taxonomy

The noun decomposes into objects. These can be classed according to their environment and key parameters. One decomposition of the nouns might be physical, mental, man made objects, and natural objects. Many of the practical things we deal with, such as an automobile, are found by descending this hierarchy.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

In modelling people, both mental and physical characteristics are of interest. Physical capacity, or how much force the body can be subjected to, might help someone designing an ejector seat, or allow checking the amount of force a person experiences during different maneuvers. Physical strength might be of interest in checking the force required to turn a knob or open a door. Thus, a person's physical modeling could be of either capacity or strength.

Mental capabilities are also of interest to the modeller. Certain tasks of abstraction may require modeling to get the right personnel fit. Also, mental endurance could be of interest for challenges like operating a vehicle for long periods of time or withstanding certain environments. Capacity and endurance are representative of these types of mental modeling.

Man made structures could be anything people make. Natural structures include things ranging from the tiniest crystal on a snowflake to our conception of the universe.

The noun, as shown in Figure 7, decomposes into the general classes of people and things.

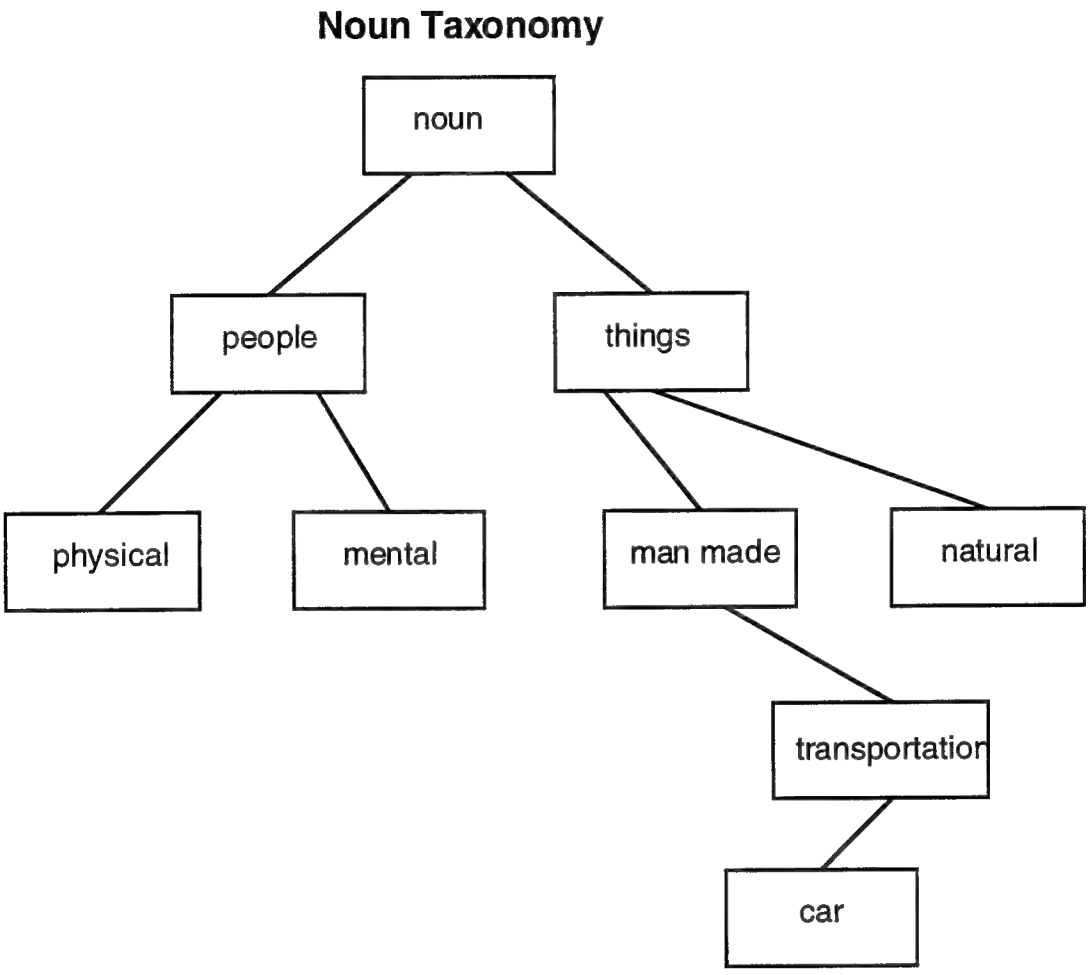


Figure 7. Noun represented in different possible classifications

The noun provides focus needed for parameterization of the verb. With the verb-noun pair, the user has a complete verb phrase. Further definition to the noun's parameters could be done with the use of modifiers. Modifiers are used to change the verb phrase's parameterization to better fit the user's conception of the process he wishes to model.

Modifier

Modifiers limit the parametric range of the verb phrase. They help the user construct a better representation of what he means. Modification also impacts the model's numerical parameterization.

Modifier Definition

The verb phrase's modifier further defines the verb phrase. Modifiers are placed directly after the noun as follows:

verb noun mod(n) mod(n-1) ... mod(1)

The user will choose the modifier(s) when defining the verb phrase representation of the model.

Each modifier adds a level of degree to the action performed. For example,

build car quickly
verb noun modifier(1)

could modify the verb build by increasing the feed rate, reducing the cycle time, and reducing the idle time. Adding another modifier,

build car very quickly
verb noun modifier(2) modifier(1)

would result in a further reduction in build parameters, possibly to their lower limit.

At this point, extension would continue according to the user's discretion.

Jerry M. Couretas

Modifier Methods

Verb phrase modification comes through increased narrowing of domain parameters. One way to present this idea would be to make each modifier a multiplier of the noun parameters that it will affect. For example, "build car quickly" could be numerically broken down as follows:

Parameters	build car		quickly		build car quickly
feed-rate	5		0.6		3.0
cycle-time	3	X	0.6	=	1.8
idle-time	1		0.6		0.6

2-14

In this example, "quickly" translates to 60% of the normal build time. We are assuming that this is a simple system where reducing each of the incoming parameters to 60% of their original value results in an equivalent reduction in build time.

Another layer of modification might be "build car very quickly." "Very" serves as an additional parameter modification as shown below:

Parameters	build car		very		quickly		build car very quickly
feed-rate	5		0.5		0.6		1.5
cycle-time	3	X	0.5	X	0.6	=	0.9
idle-time	1		0.5		0.6		0.3

The modification process could be continued on until the minimum of each parameter limitation is reached.

This is a very simple example of the modifier's ability to affect the verb phrase parameters. Extensions could be done in how the parameters are modified, when parameter modifications occur, etc.

Modifiers affect parameters in correlation with their semantic meaning. Modifier interpretation would be up to the user. How much is "very?" Or, How fast is "quickly?" In time, research might show that these modifiers are generally quantifiable over a range of domains. But, for now, this will be up to the user.

Conclusion

The system proposed here allows for complete model description in terms of a verb phrase. The verb, representing the action to be performed, accesses a discrete event model primitive. The verb's state variables would then be numerically parameterized by the noun's domain. Further control of the domain comes through the modifiers. These are placed after the verb-noun pair and modify the entire verb phrase. The verb phrase consists of a verb-noun-modifier(s) triplet that represents a discrete event model.

This representation is not complete. There are quite a few possibilities for expansion of this system. The verb phrase could be expanded to add clarification. Verb phrases could be combined for multiple action. And, the practical issues of dealing with parameters have not been addressed. The verb phrase described here is simply a way to give the user an intuitive model representation.

Using a verb phrase for model representation could open the domain of modeling to a larger group of users. We believe that the main barrier between many people and existing modeling software is their lack of computer literacy. They might use the ELI as a means of bridging this gap. Verb phrase representation could create modelers out of people who think semantically, and have no computer skills.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

A semantic representation frees the user to explore the system on the familiar grounds of natural language. A verb phrase system might open the way for brain storming, innovation and proving out of ideas before they leave the drawing board. In addition, more ideas could be checked and compared before deciding on the best option. The potential applications that might result from expanded use of discrete event modeling are wide open. And ELI could lead the way.

Analysis of Extraction and Aggregation Techniques for Model Construction in Highly Autonomous Systems

Jerry M. Couretas

Bibliography

1. Checkland, Peter, and Scholes, Jim, "Soft Systems Methodology in Action," John Wiley & Sons, New York, NY, 1990.

2. Fishwick, Paul A., "Toward an Integrated Approach to Simulation Model Engineering," Int. J General Systems, Vol. 17, 1990, pp. 1-19.

3. Fernald, James C. "English Grammar Simplified," Funk & Wagnalls Company, New York, NY, 1957.

4. Suh, Nam P. "The Principles of Design," Oxford University Press, Inc., New York, NY, 1990.

5. Zeigler, Bernard P. "Multifacetted Modelling and Discrete Event Simulation," Academic Press, London, 1984.

6. Zeigler, Bernard P. "Intelligent Agents and Endomorphic Systems," Academic Press, London, 1990.

Using Self-Organization to Develop Vector Representations of Text: A Progress Report

Ric Crabbe
Graduate Student
Computer Science Department

University of California at Los Angeles
405 Hilgard Ave.
Los Angeles, California 90024

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

Using Self-Organization to Develop Vector Representations of Text: A Progress Report

Ric Crabbe
Graduate Student
Computer Science Department
University of California at Los Angeles

Abstract

This paper presents work in progress towards a flexible and automatic method of generating representations of free text for input to neural networks. A partial algorithm to develop vector representations for words, sentences, and paragraphs with self-organization techniques is introduced. The results are analyzed, problems with the algorithm are identified, and solutions are presented.

I Introduction

One of the foremost difficulties in the creation of language processing or understanding systems is the development of the proper representation for the language in the machine. This task is complicated by the requirement that the representation be flexible enough to allow for recombination so that it may represent the infinite generative capacity of human language. This is an even more complex problem when the understanding system is a variety of neural network that requires the representation to be in the form of a vector with a reasonable number of dimensions. One possible technique for solving this problem is to use self organizing algorithms to arrange vectors representing words based upon co-occurrences of those words in a text corpus. It is hoped that this results in similar words having similar vectors, yet words can be easily combined by adding together their vectors. This paper is a report on work in progress towards the discovery of the proper self-organizing techniques needed to create this vector space, and a discussion of future directions of the work.

A Context Vectors

The peculiar set of vectors we are striving to create, sometimes called context vectors [1,4,5], are base upon these three observations:

- In large corpuses, statistics of word co-occurrence are static and a large co-occurrence factor indicates a similarity in meaning or usage. This is similar to the associational links apparent in people. Words are connected with other words and contexts.
- In high dimensional positive vector space, it is possible to create a set of vectors where all vectors are between 90° and (90 - ε)°[1] [10] apart and the number of vectors in the space is much greater than the number of dimensions of the space. Moreover, as the number of dimensions increases, the number of these nearly orthogonal vectors that can be placed in the vector space increases apparently exponentially.

[1]Where ε is small, for example 4° .

•If a vector space were large enough to support a number of nearly orthogonal vectors, when these vectors are clustered into groups based upon similarity such that the groups are nearly orthogonal to each other, there would be a large number of directions left to point in that space. Two vectors in this space could be added, creating a resultant that points in a wholly new direction in the space. This resultant would have components of the its parts, while being unique in its own right and not be involved in crosstalk with other vectors, either original or resultant.

Based upon the above observations, context vectors are defined as follows: For each distinct word in the corpus, there is an associated unit vector in some moderately high dimensional space. Before organization, these vectors are all nearly orthogonal so that the initial configuration does not influence the self-organization. After organization, the vectors are clustered according to their co-occurrence (and thus their meaning and usage) so that the clusters are far apart. When it is necessary to represent a piece of text, the vectors of the component words are summed and the resultant is normalized to a unit vector. Because the vector space is sparse, this resultant has its own individual direction while still remaining close to the vectors for its components.

B Context Vector Properties

Context vector representations share properties with many other vector text representation schemes, such as microfeatures. Most notably, both methods allocate similar vectors to similar words, combine words into sentences by some sort of summation, and have difficulty representing sentence structure [3,7,8]. Context vectors however, do have certain advantages, such as their automatic creation from example text as opposed to requiring hand coding and their having relatively fewer dimensions for a large number of words. Context vectors do represent sentences such as 'Bill hit the ball' and 'The ball hit Bill' with exactly the same code, but for some applications, this may not be important. For instance, an Air Force Intelligence neural network which recognizes scenarios for prediction or identification to a human operator, may only need to recognize generalizations about individual events [6]. The knowledge that there was a hitting and it involved both Bill and a ball may be enough for the

network to recognize the event. In fact, the loss in information may even add to the generalization powers of the network, allowing it to respond more fluidly when presented with unknown data.

II Method

The most difficult part of the context vector process is the self organization. Based on the literature, [4] it is believed that a self organizing algorithm that creates the proper vector space exists, but it is unknown to the investigators. The rest of this paper will discuss the in progress work of developing this self-organizing algorithm.

A Preprocessing

The first step in creating a set of context vectors is to develop a lexicon for the data set being used. This lexicon should contain all the words in the training set. To create the training set, a body of text is processed in two ways: first, overly common structural words that do not carry much semantics, such as articles and prepositions, are removed; then words with the same stem are grouped together as the same word and are replaced with their stem. These steps reduce the number of distinct vectors in the space and also allow for more meaningful clustering as there are more repetitions of particular word co-occurrences.

Once a lexicon has been determined, the initial vector space must be initialized so that the vectors are all nearly orthogonal. This can be done with a simple statistical method: for each element of each vector, generate a pseudo-random number between zero and one. Then evaluate a gaussian function of the generated number such as:

$$f(x) = e^{\frac{-(x-0.5)^2}{0.05}}$$

Finally, set the value of the function as a single element of the vector. When all the elements of a vector have been selected, the vector is normalized to unit length. As will be seen later, this method works extremely well. You will notice that this places all the vectors in the positive high-dimensional 'quadrant.' Without this restriction, some vectors could be as much as 180° while much

closer to other vectors. This would imply a priori relations between some words, something we hope to avoid.

Once the vector space has been initialized, the training can begin. First the text corpus is processed, removing stop words and reducing words to their stems. Then each word in the text is looked at in turn: for each word, w neighboring words on both sides are identified. The vectors associated with the neighboring words are then moved toward the vector for the word being examined. The movement is performed by moving each dimension of the mobile vector some percentage towards each corresponding dimension in the stationary vector, thus moving along a straight line between the two vectors. The resulting vector is then normalized back to unit length. The percentage that the mobile vector is moved is based on a gaussian function of how far away the two words appeared in the text, falling as the words become farther apart. The algorithm resembles the learning algorithm for self-organizing feature maps, except instead of moving the weights of neighboring nodes in the map, the vectors of neighboring words in the text are moved [5]. Neighborhoods are not a fixed function of geometry, but a changing function of the natural language text.

The above algorithm was run on two corpuses: MUC data of around 100,000 lines concerning Central and South American Politics, and a portion of the TIPSTER data, drawn from The Wall Street Journal for the years 1990, 1991, and 1992 for approximately 2 million lines. Because treating individual numbers as different words would increase the lexicon size beyond use, all numbers from the lexicon and the training data were deleted . Due to time constraints, it was not possible to perform the morphological analysis necessary to do the stemming. As a result, the lexicon for the MUC data was 17,000 words and the lexicon for the TIPSTER data was 125,000 words. The MUC data was run with a w = 4 with a gaussian of $0.1e^{(-x^2)/6}$ and later with w = 2 and a gaussian of $0.05e^{(-x^2)}$. Because of better performance, on the MUC data with the smaller window size and gaussian, the TIPSTER data was only run with w= 2 and a gaussian of $0.05e^{(-x^2)}$.

III Results

A Initialization

The random-gaussian method of initializing the vectors worked remarkably well. When initializing 17,444 vectors for the MUC data, only 200 dimensions were needed to keep the average angle between vectors at 88.83°. More interestingly, we were able to fit the 125,899 vectors for the TIPSTER data in 250 dimensions with an average angle of 89.11°. An example initialized vector is shown in Figure 1. Notice that the example vector resembles a vector from a localist representation [9]. The majority of the dimensions are close to or are zero. A single dimension is close to one, resulting in a unit vector that lies close to an axis of the space yet still far form other vectors. While it makes some intuitive sense that these vectors might turn out as they do, we are currently examining the causes and ramifications of this phenomenon.

B MUC results

After several training attempts with various size windows and gaussians on the MUC data, it was decided that a window size of 2 works best. Larger windows seem to over correlate the words. Figure 2 shows results of a single pass through the data with a window of 2 and a gaussian of $0.05e^{(-x^2)/6}$. The table displays the ten closest words to the inputted word, the farthest word from the inputted word, and their dot-products[2].

The first word, *bomb*, has a surrounding cluster that is as hoped for, at least for the closest neighbors. As we look farther away, words such as *easy* and *forget* cloud results. A less successful cluster is seen around the word *terrorist*, with the only interesting neighbors being *shining*, and *path*[3]. The final example, of *car*, shows a complete lack of meaningful organization. In all of the above examples, the word farthest away was the same, *gary*. By looking at the neighbors of *gary* and its occurrences in the text, its clear that the word just didn't appear often and remained near its position,

---

[2]Because the vectors are all unit vectors, the dot-product is equal to the cosine of the angle between the two vectors.

[3]As in the Shining Path terrorist group.

and thus orthogonal to the rest of the vectors.

word1 0.000000000412 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.011757278168 0.000000321580
0.000000000000 0.000000002987 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000002650
0.000000000000 0.000000000011 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000132188479
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000002 0.000000000000
0.000000000000 0.000000120731 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000006272 0.000000000000 0.000000000000 0.000000000000 0.016667120181
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.001872274833 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000035783072 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000125338086 0.000000000005 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000508840098 0.000000000000 0.000000000005 0.140474996410 0.000000000000 0.003404264305
0.000000583236 0.000000000000 0.000000000000 0.000000000000 0.000000324222 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000002760 0.000000000000 0.000000000000 0.000000000000 0.000000000006 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000006792703
0.000000000000 0.073607631211 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.070797570649 0.000000000000 0.000000818444 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000111913 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.041935637187 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000040100844 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000051 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000104 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000001 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000039 0.001955110817 0.000000000000
0.000000000000 0.000000008404 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000232334 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.002995837296 0.983236495581 0.000000000005 0.000000000000 0.000000000000
0.000124496394 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000
0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.000000000000 0.029657429795 0.000000000000
0.000000000051 0.000000000000 0.000000000002 0.000000000000 0.000000000000

Figure 1 - An initialized 250 dimension vector

Figure 3 shows the clusters around *terrorist* and *car* after the vectors had been trained on the data five times. *Terrorist's* cluster has improved, adding violence and attacking words as well to *shining path. Car,* on the other hand, remained in the center of a rather puzzling cluster. Further training had little effect on the clusters.

C. TIPSTER results

Because of the strange results from the MUC data, it was hoped that the broader data from the Wall St. Journal might produce less idiosyncratic findings. Figure 4 shows the clusters around the three

words *stock*, *securities*, and *cuba* after one training pass with a *w* of 2 and the gaussian: $0.05e^{(-x^2)/6}$. The most noticeable difficulty with these results is the proliferation of abbreviations. Unfortunately, do to the size of the TIPSTER data and time constraints, it was impossible to find and expand all the various abbreviations in the data. However, when the abbreviations are ignored, the clusters around *stock* and *securities* are similar to what would be hoped[4]. *Cuba*, on the other hand, has only one neighbor which could be considered to be related: *tobacco*.

Enter word: bomb
close w = exploded d = 0.999621
close w = injuries d = 0.999286
close w = destroying d = 0.999051
close w = men d = 0.999042
close w = goal d = 0.999042
close w = when d = 0.999009
close w = suffered d = 0.998970
close w = forget d = 0.998970
close w = taking d = 0.998962
close w = easy d = 0.998934
far w = gary d = 0.102042

Enter word: terrorist
close w = path d = 0.999790
close w = points d = 0.999761
close w = shining d = 0.999752
close w = described d = 0.999749
close w = summit d = 0.999747
close w = maintain d = 0.999745
close w = daniel d = 0.999744
close w = say d = 0.999743
close w = nicaraguan d = 0.999741
close w = promised d = 0.999741
far w = gary d = 0.103003

Enter word: car
close w = camilist d = 0.998083
close w = reduced d = 0.998071
close w = today d = 0.998061
close w = reaffirm d = 0.997811
close w = result d = 0.997810
close w = crisis d = 0.997773
close w = ended d = 0.997674
close w = searched d = 0.997586
close w = kidnappings d = 0.997564
close w = difference d = 0.997557
far w = gary d = 0.101287

Figure 2- Vectors trained from MUC data.

Figure 5 shows the clusters around the same 3 words after four passes through the TIPSTER data. The cluster around stock includes more words that we might hope for, although some, such as dividends, might be expected to be higher in the list. While the cluster around securities has changed, many of the words do belong in that group. As with car in the MUC trained vectors, cuba does not seem to improve with further training, but worsens.

[4]The abbreviation closest to *securities*, *scr*, is an abbreviation of *securities* .

Enter word: terrorist
close w = path d = 0.999979
close w = shining d = 0.999977
close w = violence d = 0.999973
close w = solution d = 0.999972
close w = major d = 0.999972
close w = power d = 0.999972
close w = sought d = 0.999972
close w = received d = 0.999971
close w = violent d = 0.999971
close w = attacks d = 0.999971
far w = paita d = 0.273513

Enter word: car
close w = deterioration d = 0.999881
close w = economic d = 0.999879
close w = pact d = 0.999876
close w = local d = 0.999867
close w = crisis d = 0.999865
close w = blocks d = 0.999861
close w = prevailed d = 0.999854
close w = operational d = 0.999852
close w = contents d = 0.999851
close w = month d = 0.999851
far w = paita d = 0.272950

Figure 3- MUC Vectors after 5 passes.

Enter word: stock
close w = edu d = 0.999841
close w = dividends d = 0.999840
close w = csu d = 0.999840
close w = plummeted d = 0.999838
close w = ien d = 0.999836
close w = actual d = 0.999836
close w = mnt d = 0.999831
close w = iei d = 0.999830
close w = registrations d = 0.999830
close w = stt d = 0.999830
far w = toujours d = 0.062042

Enter word: securities
close w = scr d = 0.999936
close w = statistics d = 0.999925
close w = stt d = 0.999924
close w = frx d = 0.999924
close w = noncallable d = 0.999924
close w = bankruptcies d = 0.999924
close w = registrations d = 0.999923
close w = bon d = 0.999922
close w = pricings d = 0.999922
close w = il d = 0.999922
far w = toujours d = 0.061949

Enter word: cuba
close w = forecast d = 0.998903
close w = surge d = 0.998802
close w = primary d = 0.998761
close w = lines d = 0.998751
close w = tobacco d = 0.998751
close w = indicating d = 0.998744
close w = arlington d = 0.998744
close w = harder d = 0.998742
close w = ch d = 0.998741
close w = incentives d = 0.998740
far w = toujours d = 0.061455

Figure 4 - TIPSTER trained vectors.

Despite the more general data, the performance of the system did not improve as was hoped. Figure 6, however, gives us some clues as to why this happened. In that figure, we see that the distances between some seemingly unrelated words are small, within 4°. This is an indication of the problem we encountered in the clustering: all of the vectors have collapsed so that they point in the same direction. After training, the majority of the vectors point in one direction, with some stragglers

such as *gary* or *toujours* pointing elsewhere. While the clusters of similar words were forming, the groupings were disguised by the fact that the clusters overlapped.

Enter word: stock	Enter word: securities
close w = rushed d = 0.999971	close w = scr d = 0.999992
close w = creditor d = 0.999970	close w = tax-exempt d = 0.999991
close w = actual d = 0.999969	close w = recreation d = 0.999991
close w = preferred d = 0.999969	close w = conglomerates d = 0.999990
close w = rtl d = 0.999969	close w = cgl d = 0.999990
close w = narrower d = 0.999969	close w = fis d = 0.999990
close w = represent d = 0.999968	close w = noncallable d = 0.999990
close w = edu d = 0.999968	close w = div d = 0.999990
close w = reorganization d = 0.999968	close w = elc d = 0.999990
close w = dividends d = 0.999968	close w = cut d = 0.999990
far w = toujours d = 0.077955	far w = toujours d = 0.077936

Enter word: cuba
close w = satisfaction d = 0.999768
close w = taught d = 0.999756
close w = willow d = 0.999751
close w = false d = 0.999734
close w = true d = 0.999734
close w = margins d = 0.999732
close w = cu d = 0.999731
close w = honeywell d = 0.999730
close w = realizing d = 0.999726
close w = allowance d = 0.999725
far w = toujours d = 0.077943

Figure 5 - TIPSTER trained vectors, four passes.

Enter word: money	Enter word: money	Enter word: english
Enter word: space	Enter word: castro	Enter word: banana
d = 0.999879041964	d = 0.999926479080	d = 0.997825498007

Figure 6 - Distances between unrelated words.

The reason for this collapse is not hard to identify. In the training algorithm, the vectors for two co-occurring words are moved towards each other, resulting in both vectors in a weighted average position. Since, as has been previously shown, the initial positions for the vectors are along the edges of the vector space, training tends to move the vectors toward the middle of the space. After enough iterations have been performed that the vectors begin to organize, the majority of them are pointing in the same direction in the center of the quadrant. When examining the vectors after training, most of the individual dimensions are equal, demonstrating that the vectors do point to the middle of the

surface of the quadrant.

IV. Future Experiments

At this stage in the research, there have been identified three areas of immediate future action. Two of these are relatively simple and were not already done only because of time constraints. The first is to expand abbreviations in the data so that words and their abbreviations can be treated as one and results can be more clearly understood. The second item to complete is the stemming. This morphological reduction will keep multiple versions of a word from pulling related vectors is multiple directions.

The most important future step is to solve the problem of collecting the vectors into the center of the space. One way to do this is to make the space a high dimensional sphere instead of just a quadrant, thus removing the center toward which the vectors can collapse. This however may introduce other problems as have been discussed above. It remains to be seen how large these introduced problems are.

Another solution to vector bunching problem is to change the training algorithm to prevent it. The new algorithm is obvious when the cause of the problem is considered. Instead of always moving the vector for a nearby word towards an object word, always move the more central vector towards the vector closer to the edge of the space. Since the vectors begin close to the edges along the axis, as clusters develop, they should do so along the axis and remain nearly orthogonal to each other.

V. Conclusion

The current work in progress on Context Vectors has not been entirely successful, but does show some promise. While the results are not as good as could be hoped, they do display certain aspects of what was predicted. The initialization procedure did create the nearly orthogonal vectors, and the clustering around some words did occur. Importantly, some of the major causes of the problems have been identified such as the vector collapse. Furthermore steps to rectify these problems have also been

isolated. With better preprocessing techniques such as full stemming, and new algorithms to prevent the collapse of the vectors, it is likely that the self organizing algorithm for context vectors will soon be discovered. Work on all of these fronts is currently being pursued.

VI. References

1. Caid, William R. *Context Vector-Based Text Retrieval.*

2. Caid, William R., et. al. *A Context Vector Approach to Image Indexing and Retrieval.*

3. Fodor, J. A., and Pylyshyn, Z. W. *Connectionism and Cognitive Architecture: A Critical Analysis.* Cognition, 28:1988.

4. Hect-Nielsen, Robert. *Context Vectors: General Purpose Approximate Meaning Representation Self-Organized From Raw Data.*

5. Hertz, John, et. al. Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Co., Redwood City, CA:1991.

6. Jumper, E. J. *A Multi Temporal Trainable Delay Neural Network.*

7. Lachter, J., and Bever, T. G. *The Relation Between Linguistic Structure and Associative Theories of Language Learning: A Constructive Critique of Some Connectionist Learning Models.* Cognition, 28:1988.

8. Pinker, S., and Prince, A. *On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition.* Cognition, 28:1988.

9. Shastri, L. A Connectionist Approach to Knowledge Representation and Limited Interference. Cognitive Science, 12:1988.

10. Watson, G. S. Statistics On Spheres. John Wiley & Sons, Inc., New York: 1983.

AN ASSIGNMENT BASED APPROACH TO PARALLEL-MACHINE SCHEDULING

Julie Hsu
Ph.D. Candidate
Department of Computer Science

Louisiana State University
298 Coates Hall
Baton Rouge, LA 70803

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

AN ASSIGNMENT BASED APPROACH TO PARALLEL-MACHINE SCHEDULING

Julie Hsu
Ph.D. Candidate
Department of Computer Science
Louisiana State University

Abstract

In this paper, we address the scheduling problem of allocating a set of tasks to a set of resources in order to optimize a set of objectives, eg. to minimize tardiness or maximize resource utilization, and while satisfying any given resource capacity or temporal constraints (between jobs, etc). Because this is a known NP-complete problem, we cannot in practice rely on exponential optimal algorithms to solve it. Instead, heuristic approaches are utilized which can find suboptimal solutions in a reasonable amount of time. The novelty of our approach has to do with the way allocation decisions are performed. Unlike some heuristic approaches which make only one resource allocation decision at a time, several decisions concerning multiple equivalent resources and multiple candidate operations are made at a time. Furthermore, the algorithm takes advantage of an efficient assignment algorithm which can handle balancing trade-offs between multiple conflicting objectives.

AN ASSIGNMENT BASED APPROACH TO PARALLEL-MACHINE SCHEDULING

Julie Hsu

1 Introduction

Scheduling problems have received much attention in recent years from operations research (OR) and artificial intelligence (AI) researchers, among others. We can attribute the difficulty in solving these problems to two sources of complexity. Intrinsic complexity refers to the combinatorially explosive nature of the search space while extrinsic complexity deals with the difficulty in problem formulation as well as the uncertain and dynamic nature of real-world problems [GB92].

In this paper, we are concerned with the intrinsic complexity of resource allocation problems. Because many scheduling problems have been proven to be NP-complete [GJ79], algorithms aimed at solving them may require exponential time in the worst case. Since this is undesirable for use in practical applications, heuristic approaches are adopted which yield "reasonable" solutions in a "reasonable" amount of time. Dispatch heuristics, though widely used in practice because of their computational simplicity, tend to yield relatively poor schedules. Although it is often difficult to assess the quality of a schedule because of the need to balance multiple conflicting objectives [SM94], a dispatch heuristic is unable to even address more than one objective. In addition, its locally greedy tactic ignores information inherent to the given problem that could contribute to a good overall schedule.

We describe an approach to resource allocation which is capable of balancing trade-offs in multiple optimization objectives in order to find a schedule with some guarantees as to its quality. Unlike dispatch heuristics in which only one decision is made at a time, our method can make several decisions concerning multiple identical resources and multiple eligible operations at each iteration. Moreover, knowledge about temporal constraints between activities is exploited to improve the efficacy of the algorithm.

2 The parallel-machine scheduling problem

An example of the resource allocation problem we address is shown in Figure 1. A variant of the general job shop scheduling problem, our challenge involves allocating a set of m identical machines or resources

(denoted by R_k) to a set of n activities or jobs (denoted by J_i) in order to optimize a set of objectives, eg. to minimize tardiness or maximize resource utilization, while satisfying any given capacity or temporal constraints. Machines are identical in the sense that any machine is capable of processing any job; moreover, there is a capacity constraint that no machine can process more than one job at a time. The problem description may also allow for some jobs to be excluded from being processed on certain machines. In our graph representation of the model, a potential job-to-resource assignment is denoted by a directional arc from job to resource. For example in Figure 1, job J_1 can be processed by either resource R_1 or R_3 but not by R_2. Practically speaking, it might be the case that R_2 causes bottlenecks when processing a job like J_1 and thus, we do not even want to consider R_2 as a potential resource for J_1. Therefore, the set of disjunctive arcs from a single job to a subset of the resources defines the set of candidate machines capable of processing that job. It is also the case that performance may vary among machines available to process a job. For instance, the same task may take longer to perform on one resource than another due to older model of machine, less experienced worker, etc. In our example, both J_1 and J_4 can be performed on R_3, but J_4 can be accomplished in 2 time units as compared to J_1 taking 8 time units. Since the set of machines work both independently and simultaneously, this problem has also been called problem as the parallel-machine scheduling problem [CS90]. Because of the extensive research done in factory scheduling, we will adopt its terminology, although it is clear that resource allocation problems appear in numerous real-life instances.

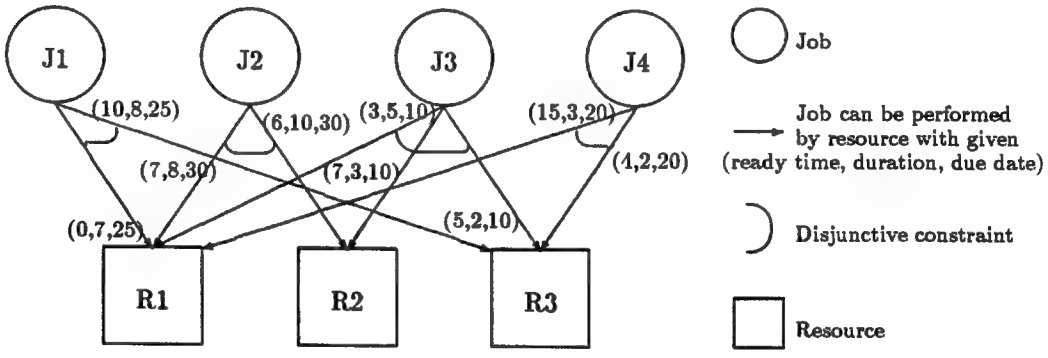


Figure 1: A parallel-machine scheduling problem

We also assume that the number of jobs is known and fixed, as is the number of machines. Since we are not concerned with uncertainty in our model, we make the assumption that machines will not break down

and are available throughout the scheduling period. Any machine setup time is assumed negligible or is included in the job's processing time. Each job consists of only one operation which cannot be preempted during processing; in other words, once a job has begun processing on a machine, it will run until completed. Because we are given the ready time (earliest start time), processing time (duration), and due date (deadline) of each job (as illustrated in Figure 2), we can utilize this knowledge to identify time precedence constraints

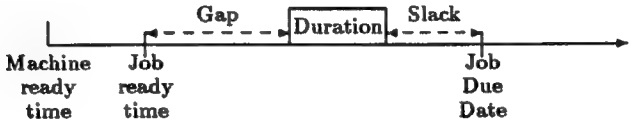


Figure 2: Timeline of a job

between jobs contending for the same resource. A job J_i precedes another job J_k on resource R_m if the earliest finish time of J_i on resource R_m (ready time plus duration) is less than or equal to the earliest start time of J_k on resource R_m. We can augment our original graph to reflect job precedences as depicted in Figure 3.

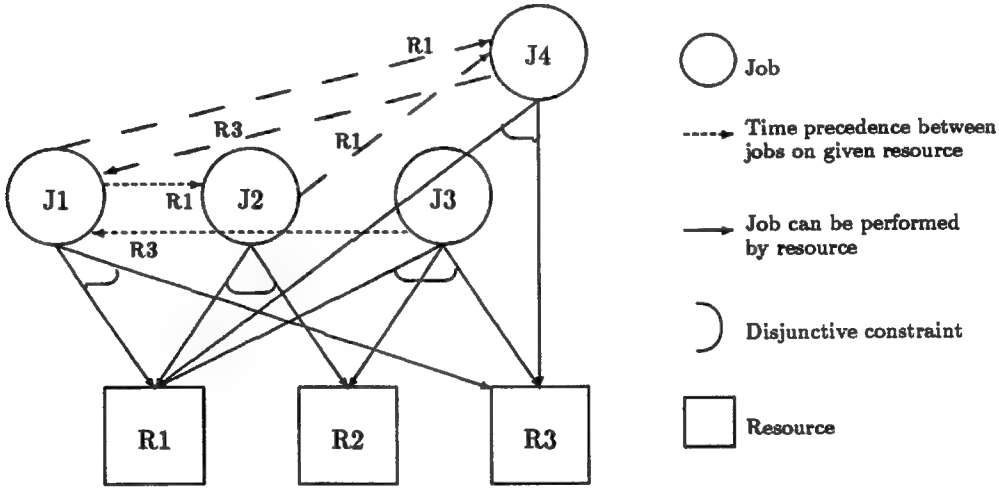


Figure 3: Scheduling problem augmented with time precedence constraints

3 The Assignment Based Algorithm (ABA)

In this section, we describe the algorithm that partitions the jobs to be scheduled into paths, such that each path corresponds to a valid sequence of operations to be performed on the same machine, each operation

has a definitive time window associated with it, i.e., a given start and finish time, and also such that all the operations are assigned to individual machines. This algorithm is inspired by previous work in job shop scheduling [Gom92] and crew scheduling in which partitioning approaches are popular ([GA88]).

A top-level perspective of the algorithm reveals two main tasks: (1) partitioning the set of jobs into levels and (2) assignment of the operations to individual machines.

3.1 Partitioning jobs into levels

We can partition the nodes of the graph in Figure 3 to more clearly reflect the precedences among jobs vying for the same resource. Once in this representation, as illustrated in Figure 4, the concept of levels becomes clear. The level of a job is the length of the longest path (number of arcs) from the resource node. In our example, although J_1 precedes J_4 for resource R_1, we do not consider transitive arcs and thus the longest path from R_1 to J_4 is of length 3, i.e., J_4 is at Level 3 for R_1.

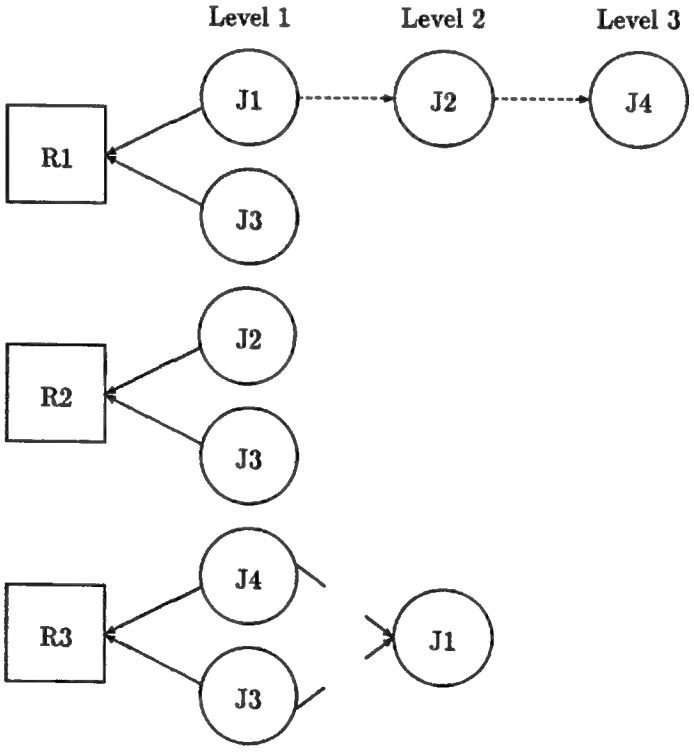


Figure 4: Partitioning candidate jobs into levels

3.2 The Assignment Problem

With the set of activities partitioned into levels, we now solve at least as many assignment problems (eg. [Chr75, GM84]) as the number of levels of the partitioned precedence graph. Figure 5 depicts an assignment problem (AP) with three node sets. Also referred to as the marriage problem, this particular case of the matching problem involves a complete bipartite graph with n node sets, i.e., n source nodes and n target nodes. Associated with each potential source-to-target pairing is a cost u_{ij}. Solving an assignment problem entails finding a one-to-one pairing of source and target nodes such that the total utility of the assignment is maximized. In terms of resource allocation, we can view source nodes as machines and target nodes as jobs we want to allocate the machines to. If the original graph is not complete, i.e., there is an imbalance in the number of source or target nodes, fictitious nodes and arcs are added to make the graph complete. In this case, any jobs assigned to "dummy" resources are considered unassigned and delayed to the next level. Likewise, any resources assigned to dummy jobs are actually not assigned a job at the current iteration.

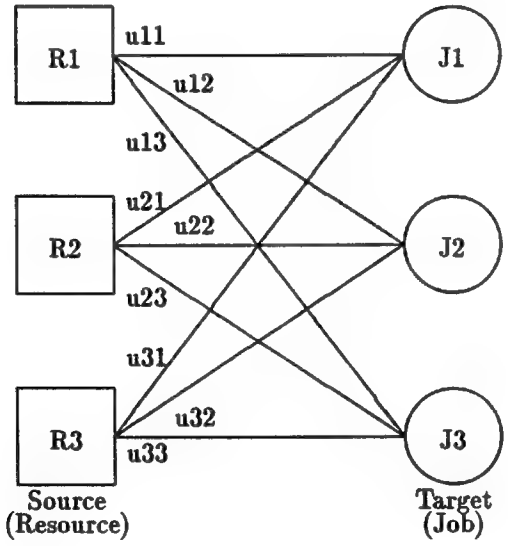


Figure 5: An assignment problem with three node sets

The power of incorporating an assignment problem algorithm (eg. the efficient Hungarian Algorithm [PS82]) as the underlying driver of our algorithm lies in two aspects. First of all, the specification of the utility function (each u_{ij} is calculated using this function) is a way to tune the system to optimize multiple goals.

In addition, since the function is simply a mathematical construct, there is much flexibility in what objectives we choose to emphasize. In our case, we experimented with several combinations of parameters and found that an aggregate function using deadline information and a combined ready time and duration measure was effective in balancing the objectives of minimizing tardiness and minimizing makespan. Secondly, each solution to an assignment problem results in up to the total number of the resources being allocated to jobs, except in the case where there are no candidate jobs for a particular resource.

3.3 ABA

We can now combine our two top-level tasks to form the Assignment Based Algorithm. After partitioning candidate jobs into levels for each resource, the method iterates through each level, solving an assignment problem with the resource set and the candidate jobs at the current level. If all resources have at least one candidate job, then all resources are guaranteed to be allocated in this iteration. Any leftover jobs not scheduled (or scheduled to dummy resources) in this iteration are added to the candidate jobs at the next level. This process continues until all jobs have been assigned to individual machines. Figure 6 summarizes ABA.

```
Given a parallel-machine scheduling problem in which n jobs are to be scheduled on m resources:
    Partition the jobs into levels according to time precedence constraints
    Begin Loop
        If the precedence graph is empty, terminate and return the schedule
        Else,
            Solve an assignment problem with the resource set and the current
             set of candidate jobs, i.e., jobs at the current level
            Update the precedence graph by removing all scheduled jobs and
             propagating constraints, i.e., delay any unscheduled jobs to the
             next level and then update the level
    End Loop
```

Figure 6: The Assignment Based Algorithm

A simplified variation of ABA is possible in which, at each iteration, all jobs are candidates to be assigned to the machines. Any jobs not assigned at the current iteration are candidates in the next iteration. Because the

jobs are not organized into levels, there is no need to generate the partitioned precedence graph. Experiments with ABA-II were conducted and are discussed in the next section.

4 Experimental Results

4.1 Generation of test set

In order to evaluate the performance of ABA, we used a test set of 60 randomly generated scheduling problems, similar in nature to the Constraint Satisfaction Scheduling benchmark proposed in [Sad91]. In order to simulate realistic scheduling scenarios, we varied certain conditions of the problems: 3 deadline ranges and 2 resource sets. The number of jobs was fixed at 80 with the number of resources fixed at either 4 or 8, the idea being that the smaller the number of available resources, the increased chance of bottlenecks and hence a larger makespan (completion time of the last job on all machines). The deadlines of the jobs are considered tight (T), median (M), or wide (W); the tighter a job's due date, the greater the chance that the job will be tardy. Interleaving these two parameters yields 6 categories of scheduling problems: T4, i.e., tight with 4 resources, T8, M4, M8, W4, and W8. For each category, 10 problems were randomly generated yielding a total of 60 test cases.

4.2 Performance criteria

[CS90] identify three classes of performance criteria: completion time based, due date based, and flowtime based measures. Completion time based criteria involve the makespan, which is important when under the assumption that the system cannot be shut down until all resources have finished processing. Due date based measures evaluate the overall tardiness, the total number of tardy jobs, etc. These are usually quite important factors since there is often a cost associated with delayed production and deliverables. Of less importance in some domains are the flowtime based measures which look at resource utilization and idleness.

4.3 Dispatch heuristics

Dispatch heuristics are often employed in real-world applications because of their computational simplicity. The set of jobs are first sorted according to the chosen dispatch rule; eg. with the earliest due date first heuristic (EDD), activities are sorted by due date in ascending order. A schedule would then be created by "dispatching" the jobs one at a time until all jobs are assigned to individual machines. Unfortunately, it is quite likely that the resulting schedule is "poor" in some sense; for instance, executing the schedule

would yield more than an acceptable number of tardy jobs. Unlike methods which employ search techniques, there is no backtracking, i.e., once an allocation is made, it is not considered again. Dispatch heuristics are also incapable of utilizing knowledge about inter-activity relationships such as time precedence constraints. Furthermore, a dispatch heuristic is based on only a single objective, such as to minimize slack, which it cannot even address well because of its locally greedy tactic.

4.4 Discussion

With the generated test set, we compared ABA against several popular dispatch heuristics as well as its own variation ABA-II. The dispatch heuristics include Earliest Due Date (EDD), Minimum Slack (MinSlack), First Come First Served (FCFS), and Shortest Processing Time (SPT). The percent reduction in makespan, tardiness, number of tardy jobs, and idleness using ABA versus each of the other dispatch heuristics and also ABA-II were calculated and plotted (see Appendix). The results of all 60 test problems are interpreted in the following manner: positive results indicate that ABA created a schedule with a better value with respect to the corresponding performance criteria. For instance, looking at the Makespan Comparisons, ABA outperformed MinSlack, EDD, and FCFS on all 60 test cases. There were a few problems in which SPT generated schedules with smaller makespan values, and some in which they were equivalent. For the Tardiness Comparison, it is worth noting that compared to MinSlack and FCFS, ABA was more effective at reducing the tardiness the tighter the deadline range (Problem Set 1 having the tightest deadline range, increasing to Problem Set 6). The First Come First Served heuristic outperforms all other heuristics and ABA when looking at resource utilization or the total idle time of the machines. But since we are usually concerned with balancing multiple conflicting objectives and FCFS is not effective at reducing the makespan or tardiness, FCFS still yields relatively poor schedules. It is also interesting to note that ABA did not always outperform ABA-II, which confirms that the quality of the resulting schedules can be attributed to the incorporation of the assignment algorithm and that the partitioning scheme is more of a contribution to efficiency.

5 Conclusions

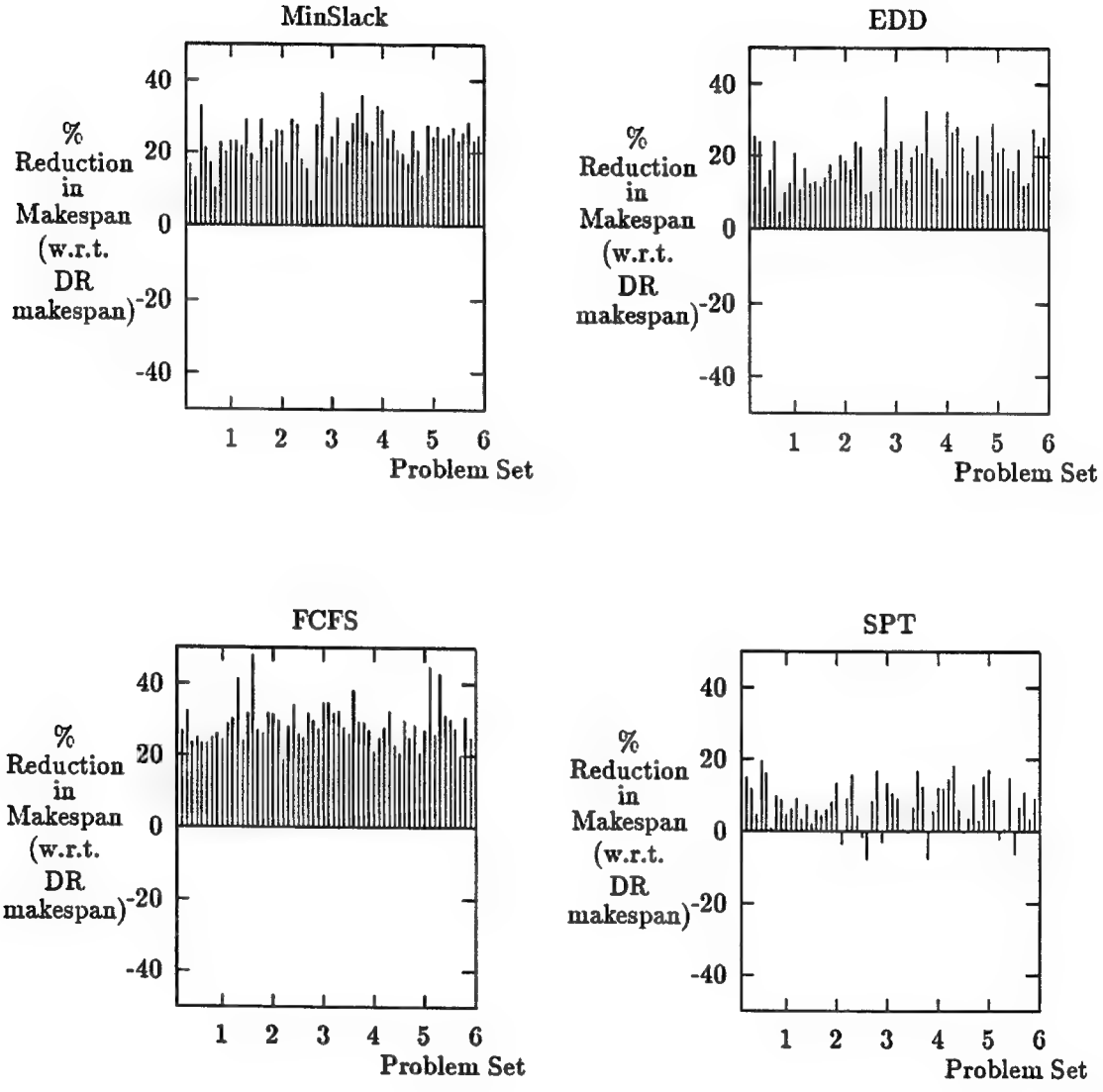
In this paper, we have presented a method for allocating a set of identical resources to a set of jobs which creates higher quality schedules than those generated by dispatch heuristics, yet still without the undesirable consequences of exponential algorithms. The approach takes advantage of an efficient assignment algorithm capable of making several allocation decisions at a time. Furthermore, the use of the manipulable objective

function allows us to optimize multiple conflicting goals. And by exploiting knowledge about time precedences among jobs contending for a resource, we consider only what is necessary at each iteration. The algorithm can be incorporated into a system responsible for maintaining a resource-based perspective, i.e., one that focuses attention on the analysis of resource contention, and its objective function can be tuned to optimize whatever parameter(s) are deemed important to the quality of the resulting schedule.

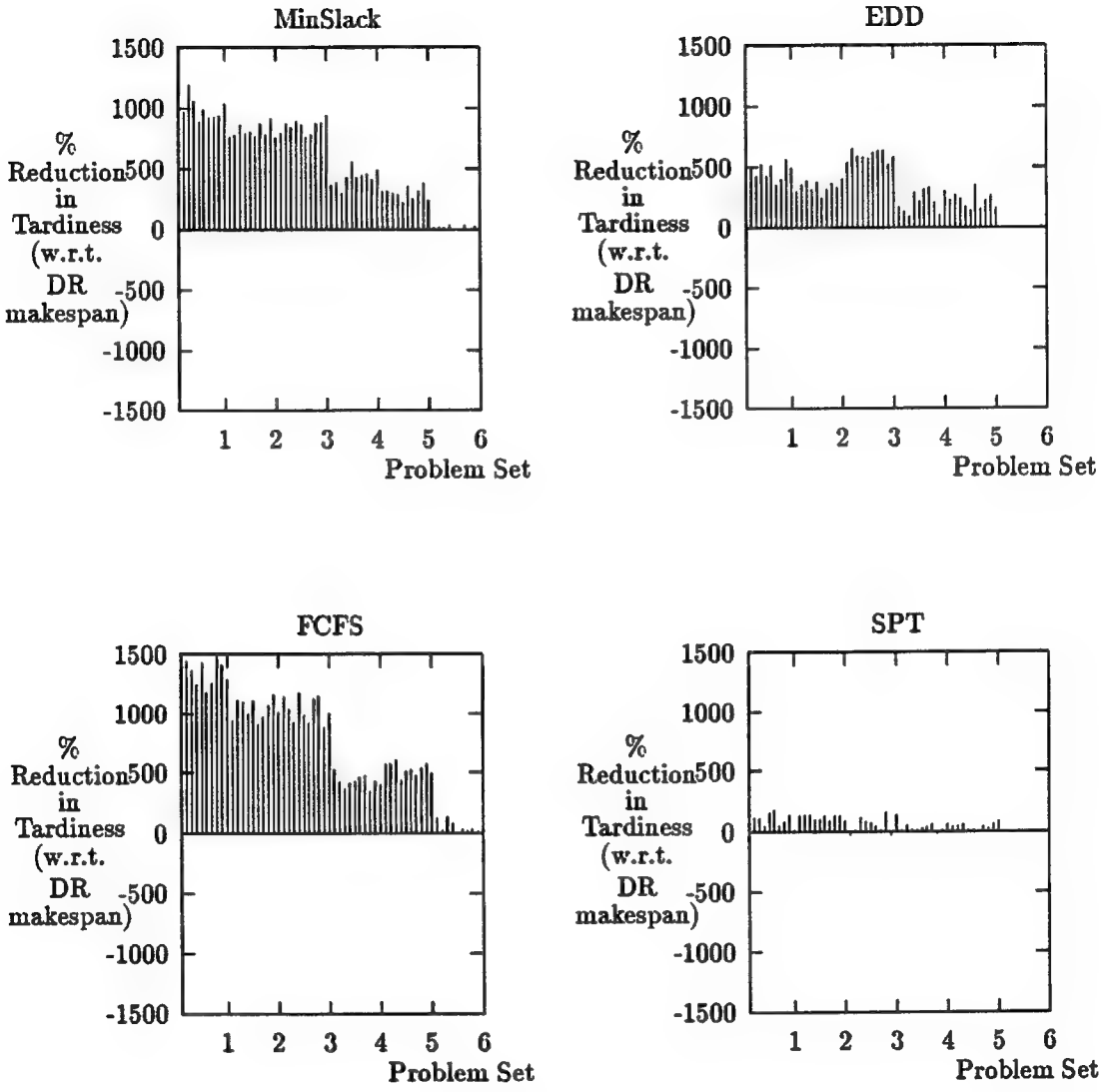
6 Appendix

Each of the following graphs shows the results of comparing ABA to the indicated dispatch rule (DR) for the 60 scheduling problems in the test set. The 60 problems are categorized as follows: Problems #1-10 (Test Set #1) is of type T4 (tight deadline range with 4 resources), problems #11-20 (Test Set #2) is of type T8 (tight with 8 resources), and so forth for categories M4 (median-4), M8 (median-8), W4 (wide-4), and W8 (wide-8).

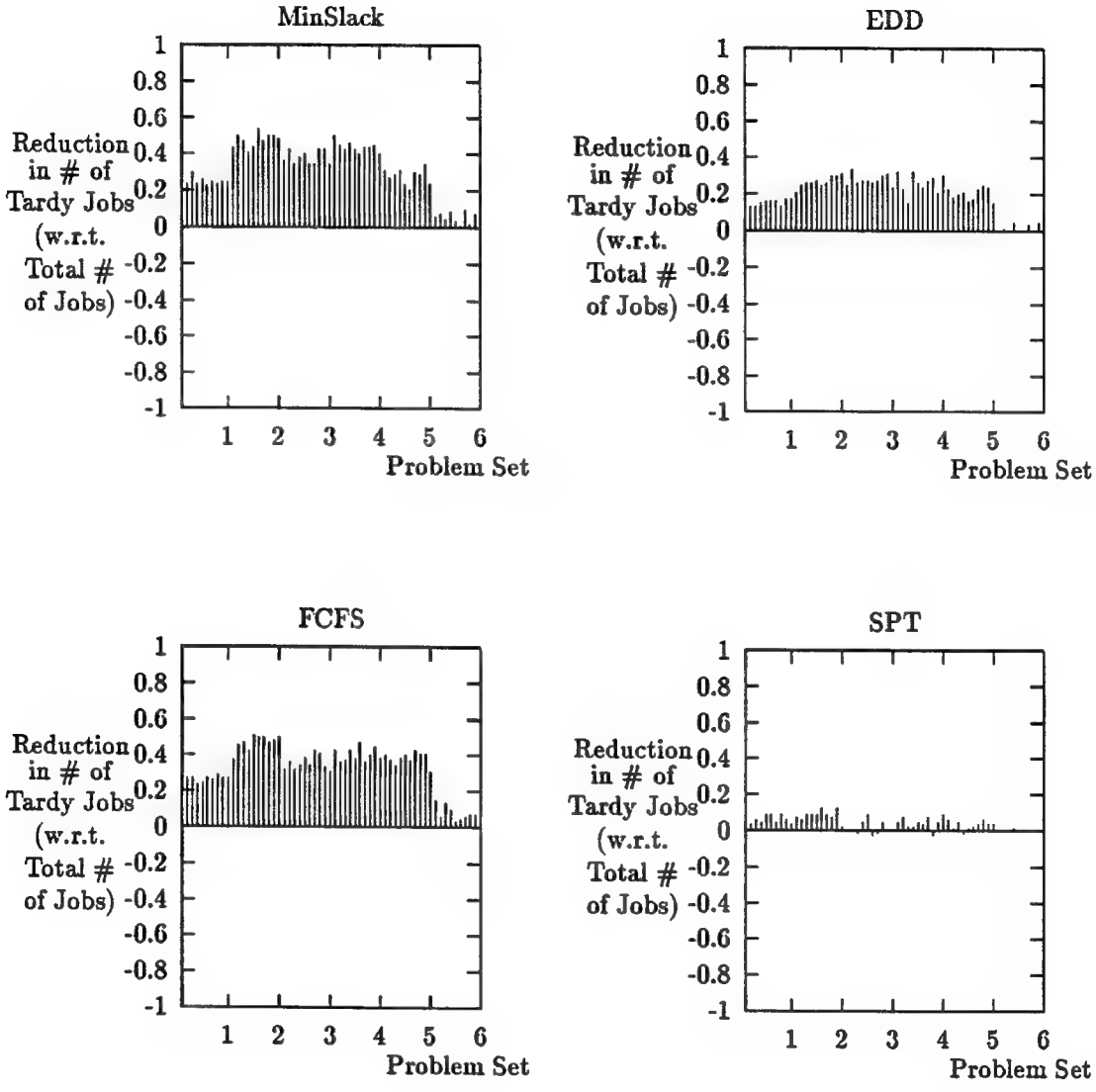
Makespan (Completion Time) Comparison



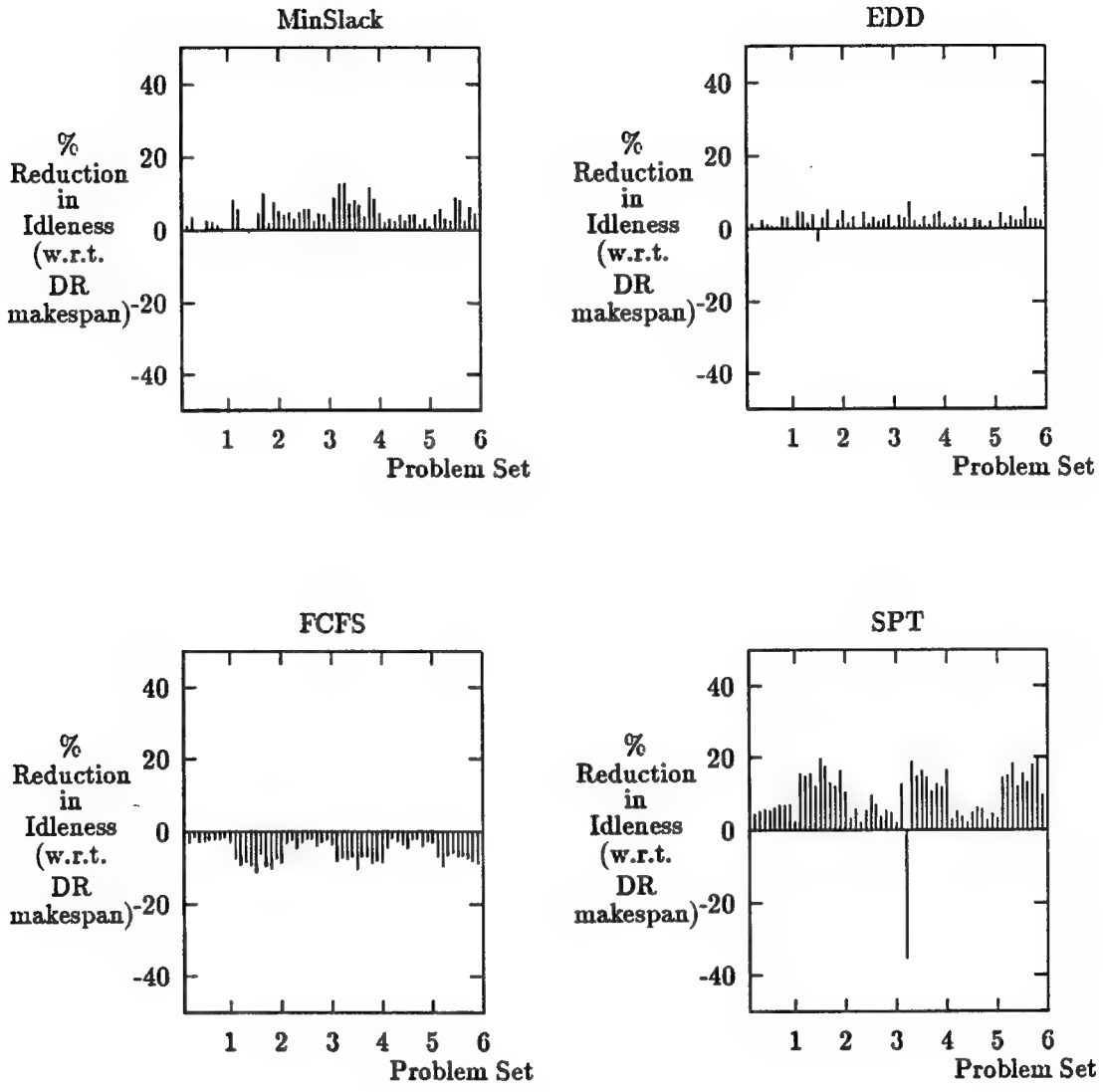
Tardiness Comparison



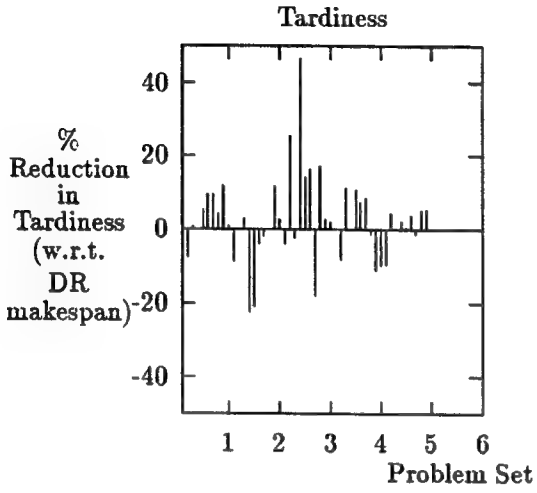
Comparison of Number of Tardy Jobs

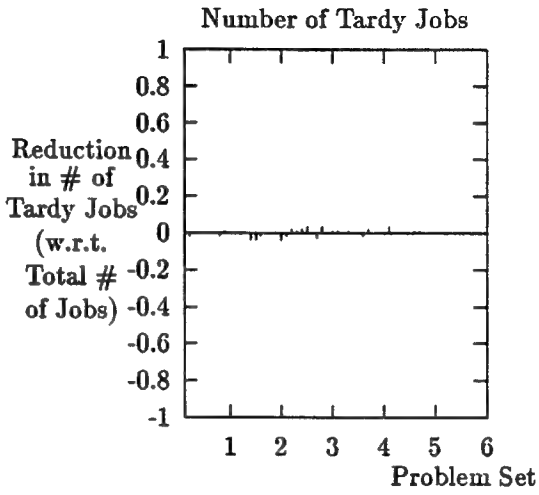


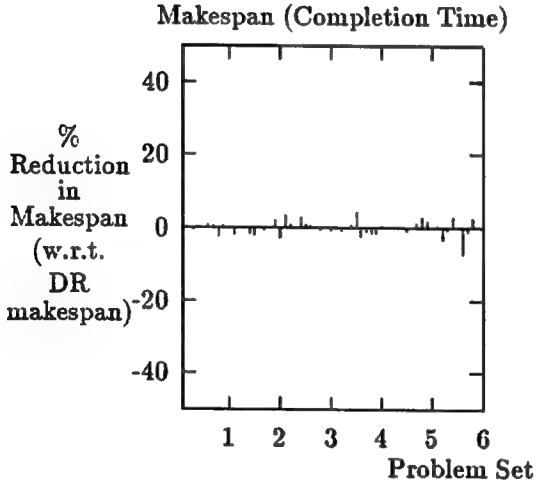
Resource Utilization / Idleness Comparison

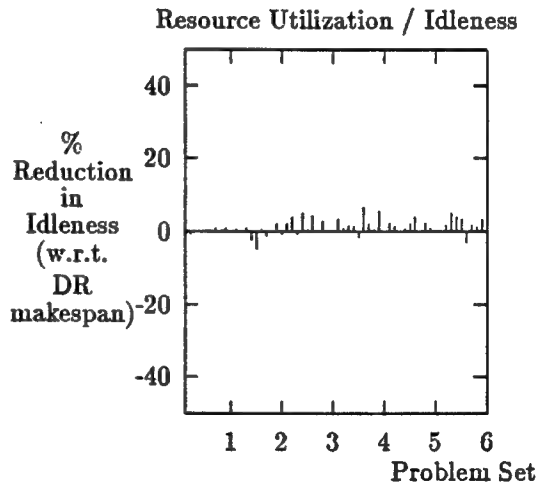


Comparison to ABA-II









References

[Chr75] N. Christofides. *Graph Theory: An Algorithmic Approach.* Academic Press, 1975.

[CS90] T.C.E. Cheng and C.C.S. Sin. A state-of-the-art review of parallel-machine scheduling research. *European Journal of Operational Research*, 47:271–292, 1990.

[GA88] C.P. Gomes and M.T. Almeida. Pairing generation - a graph partitioning approach to a short haul fleet problem. In *Proceedings of AGIFORS Group Meeting*, Copenhagen, 1988.

[GB92] C.P. Gomes and H. Beck. Synchronous and asynchronous factory scheduling. *Journal of the Singapore Computer Society*, 5(2), 1992.

[GJ79] M.R. Garey and D.S. Johnson. *Computers and Intractability.* W.H. Freeman and Co., San Francisco, 1979.

[GM84] M. Gondran and M. Minoux. *Graphs and Algorithms.* John Wiley & Sons, 1984.

[Gom92] C.P. Gomes. Achieving global coherence by exploiting conflict: A distributed framework for job shop scheduling. Unpublished Ph.D. thesis, University of Edinburgh, 1992.

[PS82] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization - Algorithms and Complexity.* Prentice-Hall, Inc., 1982.

[Sad91] N. Sadeh. *Look-ahead Techniques for Micro-opportunistic Job Shop Scheduling.* PhD thesis, Carnegie-Mellon University, 1991.

[SM94] K.P. Sycara and K. Miyashita. Evaluation and improvement of schedules according to interactively acquired user-defined criteria. In *Proceedings of ARPA/Rome Lab Planning Initiative Workshop*, pages 155–164, 1994.

AUTOMATIC EXTRACTION OF DRAINAGE NETWORK FROM DIGITAL TERRAIN ELEVATION DATA (DTED)

Andrew J. Laffely and Mohamad T. Musavi
Graduate Student Associate Professor
Department of Electrical & Computer Engineering

University of Maine
5708 Barrows Hall
Orono, ME 04469-5708
Phone: (207) 581-2243
Fax: (207) 581-2220
E-mail: musavi@eece.maine.edu
E-mail: alaff@eece.maine.edu

Final report for:
Summer Faculty Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC

and

Rome Laboratory

July 1994

AUTOMATIC EXTRACTION OF DRAINAGE NETWORK FROM DIGITAL TERRAIN ELEVATION DATA

Andrew J. Laffely and Mohamad T. Musavi
Graduate Student Associate Professor
Department of Electrical & Computer Engineering
University of Maine

Abstract

In this report a straight forward approach for automatic extraction of drainage network from Digital Terrain Elevation Data (DTED) of the Defense Mapping Agency (DMA) is presented. The approach is based on common image processing methods such as frequency domain filtering, spatial domain histogram equalization, binarization, thinning and other techniques. The approach has been examined with several DTED files. The results obtained from this approach are as good as, or better than, other methods. The approach was developed using Khoros image processing system and C programming. Khoros has excellent visually based environment and its rather extensive library of functions helped avoid writing of commonly used codes.

AUTOMATIC EXTRACTION OF DRAINAGE NETWORK FROM DIGITAL TERRAIN ELEVATION DATA

Andrew J. Laffely and Mohamad T. Musavi

I. INTRODUCTION

A. Techniques

Digital elevation data files are important sources of information that have been used in many remote sensing applications such as geometric rectification [Wiesel, 85], terrain related radiometric correction [Colby, 91], image registration [Dubayah & Dozier, 86], topographic map update [Gugan & Dowman, 88], and extraction of such important features as drainage networks [Argialas, Rouge, & Mintzer, 88; Jenson & Domingue, 88; Hardipriono, Lyon, Li, & Argialas, 90].

Drainage networks have been applied in water resource assessment, hydrogeometric analysis, interpretation of soils and rocks, and segmentation of textures and structures. In principle the task is straight forward. Water drains from high to low elevation along channels that are connected to one another according to well-defined rules [Parvis, 50; Howard, 67; Argialas, 85]. Majority of the techniques used for extraction of water drainage are based on low level pixel (cell) labeling by using local operators and subsequent pixel grouping according to local criteria [O'Callaghan & Mark, 84; Jenson, 84]. Recently, Qian, Ehrich, & Campbell [90] applied both local operators and global reasoning to improve the performance of previous results. Water drainage extraction techniques normally assume one pixel wide networks that may cause difficulty in flat areas with flood patterns [Lee, Snyder, & Fisher, 92].

The technique presented here is based on two fundamental steps. In the first step, a complete and connected candidate for a possible drainage network is found. In the second step, the possible drainage network is pruned to remove unnecessary links that are not part of the drainage network. The first step is based on a set of image processing routines such as data preprocessing, frequency domain filtering, spatial domain histogram equalization, binarization, thinning and image blending. The Khoros image processing environment [The University of New Mexico, 1990] has been used in this study because its rather extensive library of functions helped avoid writing of common-used codes.

B. Databases

There are two common sources of elevation data, the Defense Mapping Agency (DMA) Digital Terrain Elevation Data (DTED) and the United States Geological Survey (USGS) Digital Elevation Model (DEM) data. While the experiments of the previous studies were all conducted on the DEM data, those of this study used DTED database. This database consists of many data files of digital terrain elevation data. Each data file is a cell defined by a matrix with longitude and latitude coordinates (x, y) and elevations (z). A cell covers a 1 degree by 1 degree geographical reference area. The sampling interval for elevation data is 3 seconds by 3 seconds. Absolute horizontal and vertical accuracy of DTED data are 50 and 30 meters respectively; while, relative accuracys within a cell are 30 and 20 meters [Military Specification, 86]. DTED data is available on CDROM and 9 track 1/2" magnetic tape.

II. METHODOLOGY

The following sections present a step-by-step description of the proposed methodology for automatic extraction of drainage network. A block diagram of the operations is shown in Figure 1. In order to show the result of each operation, a sample DMA data file has been selected. The file is a matrix of 128x128 elevation data. A three dimensional plot of this elevation file is shown in Figure 2. The elevation in this area changes from 420 to 4160 meters as indicated in the plot. For the sake of visual representation of the data and the future results, the elevation data is converted to gray level representations from 0, for the lowest elevation, to 255 for the highest elevation. The gray level representation of elevation data of Figure 2 is shown in Figure 3. Note that all future operations are performed on the original elevation data and only for demonstration purposes the results are converted to gray level images. In Figure 3 the lower elevations are almost featureless because of the great contrast of elevation present in this data.

A. Data Conversion

A DTED file supplied on CDROM is a 1201x1201 matrix of elevation data with the DMA data format. This data is first converted to the format accepted by Khoros image processing software and then is split into smaller matrices.

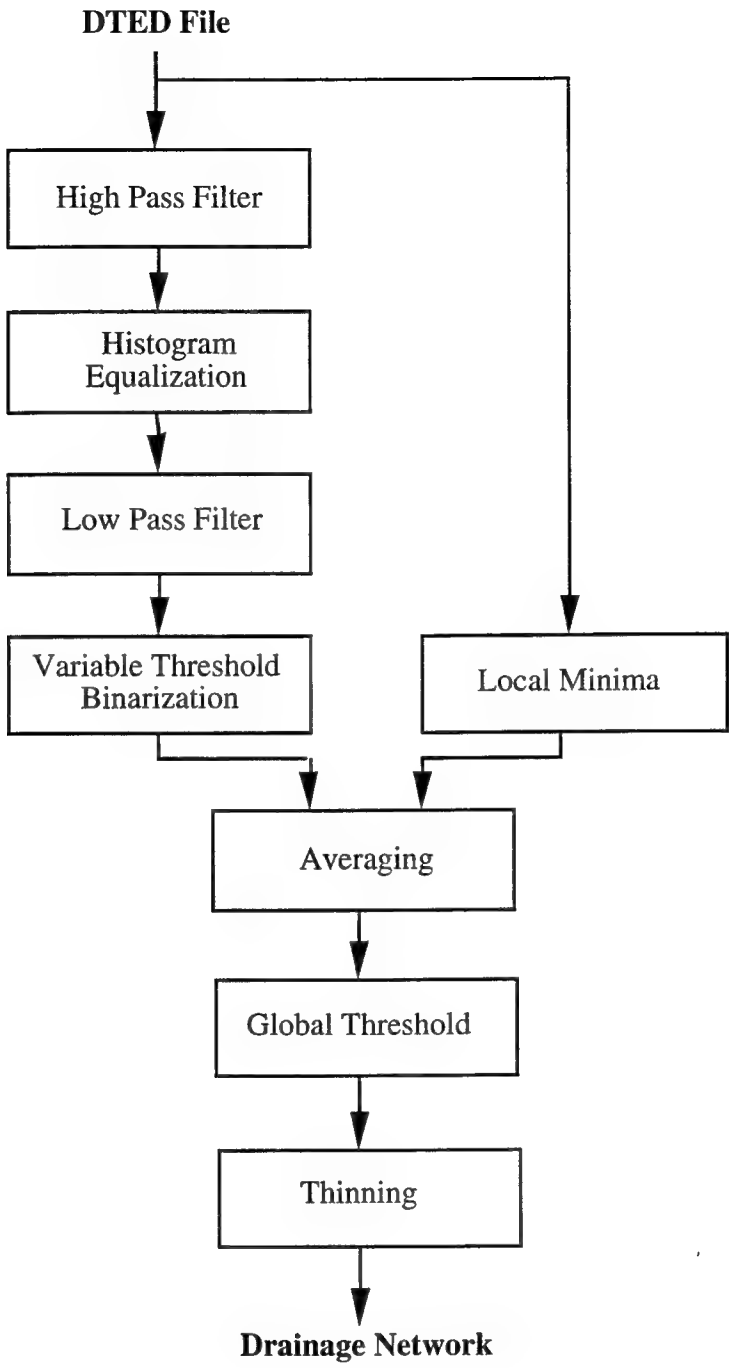


Figure 1. A block diagram for the automatic extraction of drainage networks.

B. High Pass Filter

The first step of the data processing is to bring up important features such as peaks and valleys that are necessary for identification of water drainage. A common tool is the application of a high pass filter (HPF). This can be performed both in spatial domain and frequency domain; however, frequency domain operation is more accurate and provides more flexibility for changing the cut-off frequency of the HPF. The cut-off frequency proves to be an important factor in handling different geographical areas. For example, in an area with high elevation changes the cut-off frequency can be quite low. While in an area with low elevation changes the cut-off frequency has to be higher in order to identify valleys with smooth changes that might be part of a drainage network or flood pattern.

To perform a HPF in frequency domain the data is first submitted to a Fast Fourier Transform (FFT) program to convert the spatial domain data to the frequency domain data. The data is then passed through the characteristic function of the HPF and then to an inverse FFT to recover the processed spatial data. Figure 4 shows the result of a high pass filter on the test data. The HPF was a first order exponential filter with a normalized cut-off frequency of 0.1.

C. Histogram Equalization

Although a possible drainage network might be recognizable by human eye in Figure 4, the data still poses difficulties in automatic extraction by computer. To further enhance the possible drainage network, the histogram of the data is equalized using a histogram equalization program. This step further enhances the drainage networks as shown in Figure 5. A comparison of Figure 5 and 3 shows the significance of the histogram equalization operation.

D. Low Pass Filter

Along with the enhancement of the high frequency and histogram information, the noise in the data is also enhanced. To suppress the noise, a low pass filter (LPF) is used. The result of LPF on the equalized data is shown in Figure 6. The filter used in this example was a second order Butterworth filter with the normalized cut-off frequency of 0.2. This provides an overall band-pass effect on the input data. Note that the process of implementing LPF is similar to that of HPF explained in section B above.

E. Binarization

The elevation data can now be segmented for isolation of the possible drainage area by a binarization process. Using a fixed threshold for the entire data file will cause either loss of drainage data or introduction of excessive points, depending on the value of the threshold. Therefore, a variable threshold binarization [Musavi, Shirvaikar, Ramanathan, & Nekovei, 88] was used. There are several sub-steps that are taken to accomplish this binarization. The threshold for an elevation data point *n* is defined as,

$$T_n = K_n * STDDEV_n + MEAN_n \qquad n = 1, 2, ..., M \tag{1}$$

M is the total number of data points in the file. K_n, $STDDEV_n$ (standard deviation), and $MEAN_n$ (mean) are defined as,

$$K_n = a * HIST(n)/M \tag{2}$$

$$MEAN_n = \Sigma P_i/(2*d+1)^2, \qquad i = 0, 1, 2, ..., (2*d+1)^2 - 1 \tag{3}$$

$$STDDEV_n = [\Sigma(P_i - MEAN_n)^2/(2*d+1)^2]^{1/2}, \tag{4}$$

Where *a* is a constant (3 in our experiments). *HIST(n)* is the histogram element corresponding to the n-th data point, P_i is the value of the i-th data point in a local area around the n-th point, and *d* determines the size of the local area. For example, d=1 for a 3x3 area, d=2 for a 5x5 area, and so on. The result of the variable threshold binarization is shown in Figure 7. One option for drainage network extraction was to apply a thinning algorithm at this stage, the result of which was a fairly well connected set of paths. However, the thinning algorithm output might not accurately provide the drainage paths for all areas. This is especially true where a given path is bordered on one side by sharp elevation changes and on the other by more gradual ones.

F. Local Minima

The above binarization process provides a general area of the possible drainage network. The local minima in this general area provides a more accurate representation of the drainage network. One should note that the local minima alone will not be sufficient to provide a connected drainage network due to its local reach. The significance of the generalized drainage area is in its global nature, global in a sense that it provides a connected path for the network. Therefore, a merger between the general area and the local minima will provide a much better data set for extraction of the actual drainage network.

A simple routine is used to find the local minima in both the x, y, and both diagonal directions from the original data. Each data point in the original elevation file is compared to the four that precede it and the four that follow it (9 points in each local array). If the elevation of the test point is lower or equal to all the points of the local area, a corresponding point in a binary representation file is set to 1. If any one of the points in the local array is lower than the test point, then a 0 will be recorded for the test point. This process is repeated for all rows, columns and diagonals of the data file. As a result, any point in the output file can be set based on only one testing direction. This enables the extraction of drainage paths in the four connected neighbors of the test pixel. Figure 8 shows a binary file indicating the local minima of the original elevation data. Notice in Figure 8 that some points where the drainage paths seem to converge are not set. This is a direct result of the local minima algorithm and part of the reason that it alone is unable to extract a connected drainage path.

G. Local Blending and Global Thresholding

After the two binary files are created, the local minima and the variable threshold binarization outputs, they need to be combined in a manner that enhances the connectivity and reduces noise. First, a weighted sum of the two files is used to combine the information into a single image. The optimization of the weighting factor has not been extensively studied but adding one part variable threshold information and two parts local minima data, has given consistent results. In other words, after the summation a 4 level (0, 1, 2, and 3) data file, as shown in Figure 9, is created. Note that the local minima helps to identify the accurate location of the drainage network in the general area of drainage.

Considering only the local minimum points of Figure 8, one can identify a possible drainage network that is not completely connected. However, the information in the general drainage area can be used to create the appropriate connectivity. One way to do this is by a local blending or smoothing operator on the 4 level data file created above. In the experiments reported here two identical smoothing operators, 7x7 averaging masks, were applied in sequence to the 4 level data file. The result of this operation is shown in Figure 10. This operation closes small gaps in the local minima drainage network while reducing noise. After smoothing, a global thresholding, set at 1.5, is applied to successfully extract a possible drainage network, Figure 11.

H. Thinning

The output of section G above is a possible drainage network with more than one pixel in width. A thinning algorithm can be applied to find a one pixel wide network, Figure 12. At this

stage there might exist certain isolated pixels or segments that are created as the results of thinning algorithm. These sporadic pixels can be removed by a simple length thresholding filter. The thinned drainage network of Figure 12 has been overlaid on the original data in Figure 13.

I. Direction of Flow

With a nearly complete drainage network, several other steps of processing may be done to extract more information from the DTED file. One of these is the drainage flow direction. Some approaches to this have been investigated as part of this program, but complete methods and code have not been developed due to time limitation. The next several paragraphs contain this approach.

The input to this stage of processing could come from one of three different steps described earlier.

First, the thinned drainage network of part H could be used. This being the final result of the tested system thus far makes it the obvious choice, but the problems with the Khoros thinning algorithm especially the connectivity ones could create problems in finding the direction of flow. If a new thinning algorithm were developed this would be an excellent place to begin.

Second, the wide drainage network of part G could be used. To do this, the direction of flow algorithm could be given two tasks. One, to find the flow direction of each pixel in the wide network and two, to use this information to thin the resulting network where needed. The result of this may give much more insight than the one pixel drainage system of step H, by actually allowing rivers to have a width. This process will be discussed further.

Third, if a wide network can be used and then thinned based on direction of flow, it is conceivable to use the output of the variable threshold binarization as a starting point. This would cut several steps of processing including the time consuming smoothing steps.

At this point the direction of flow algorithm will be described in terms of the second input option, however the methods presented here could be transformed to meet the demands of the other two input cases. The input image in all three cases is binary with one level equal to the value 1, and the zero level is 0.

1) The first step of this approach is to multiply the binary drainage network file by the original DTED file pixel by pixel. This will weight each pixel in the drainage network by its elevation value.

2) Create a mask image consisting of the connectivity of the above network. In this file all none zero pixels contained in a connected drainage path will be assigned a numerated cluster. The number of clusters will be the number of individual drainage networks in the test area. This step is not too difficult to construct and is almost complete. It involves searching the input file pixel by pixel till a non-zero pixel is reached; the location of this is then stored. Once the first non-

zero pixel is reached all other non-zero pixels that are connected to it, be it directly as a neighbor pixel or through a complicated string of non-zero pixels are stored in the connection mask and assigned to a cluster. After all the pixels for that network have been identified, a file search for another non-zero pixel begins one pixel after the location of the first pixel in the last cluster. This time the search will not only look for non-zero pixels but also check whether the pixel has been assigned to a cluster to avoid repetition of clustering. The search will continue to the last pixel in the file assuring that all possible networks have been found.

3) With the clustering of networks a length threshold can easily be used to remove insignificant or error drainage paths.

4) Create a pixel by pixel direction mask by, comparing the elevation of each pixel with its immediate neighbors and selecting the neighbor pixel with the greatest absolute difference as the direction. A more elaborate method is to take a weighted average of all possible directions, but this is more complicate and may not be necessary.

5) Perform thinning and path check in one (iterative) step. Test each direction against the input file. If the direction of the test pixel points from a zero pixel on one side to a non-zero pixel on the other, delete it from the network. If the direction of the test pixel points from a non-zero pixel to a zero pixel, change the non-zero pixel to 1. If the direction of the test pixel points from one non-zero pixel to another, do nothing. This procedure will delete some of the network at the higher elevations, but in those regions the drainage network will not be visible in most other mapping media. Clustering and directions will need to be recalculated after each iteration. It is uncertain what number of iterations will be sufficient to obtain the correct network.

6) Use the pixel directions to obtain a global sense of drainage direction. This will be done by averaging the directions along each path.

This approach is untested and may need to be revised.

J. Drainage Vectorizatoin

A complete network or even a partial network can be represented as a set of connected or disconnected vectors. Each network, of sufficient length, can be simplified to a set of vectors that follow its path. The resulting set of vectors may be useful in registration of satellite imagery that has a perpendicular frame of reference. The relative vector lengths and angles should be similar in the DTED and imagery allowing the registration process to overcome rotational, translational and scaling problems.

III. CONCLUSION

A straight forward technique has been presented for automatic extraction of drainage network from Digital Terrain Elevation Data (DTED) of the DMA. It has been shown that an appropriately assembled set of commonly used image processing techniques is able to extract drainage networks from elevation data. However, there are other issues that need to be addressed before the proposed methodology can be adopted as a robust and reliable tool for the task. Among these issues are determination of direction of flow, test of the methodology with elevation data from different geographical areas, test of the accuracy by comparing the extracted drainage networks with ground truth data, and refinement and optimization of the parameters used in the methodology. A neural network approach could have also been investigated and devised for the task, however, due to lack of training data, the implementation was not feasible at this stage. In addition to many useful application of drainage networks, they can also be used for automatic registration of unregistered images because drainage networks are unique features of elevation data that can be used effectively for registration purposes. Expansion, modification, and utilization of the developed methodology will be addressed in future research proposals.

ACKNOWLEDGMENT

We would like to give our thanks and appreciation to all those who helped us in the completion of this research project.

First, we would like to thank our sponsors, the Air Force Office of Scientific Research and the Research and Development Laboratories. Special thanks go to Mr. Garry Barringer, Technical Director, Rome Laboratory Directorate of Intelligence and Reconnaissance (RL/IR), and Mr. Joseph Palermo, Chief, Image System Divisions, for providing us with the opportunity to work on a very interesting project.

We would especially like to thank Mr. James McNeely, our focal point, for providing the project and helping us to understand and conceptualize it. Without his efforts we would not have been able to successfully accomplish the goals of this research.

Our thanks are also extended to Mrs. Delores Spado for giving us administrative support, to Mr. Jeff Hanson for providing us with a Macintosh PowerBook for writing this report, and to Mr. James Siefert and Mr. Ron Woerner for the ADRI images and the DTED data files.

Finally, we would give our special thanks to Mr. John Pirog for providing us with a powerful computing facility and to Mr. Mike Scully for installing Khoros software and assisting us with many other miscellaneous computing tasks.

REFERENCES

Argialas, D.P. (1985). A structural approach towards drainage pattern recognition, Ph.D. Dissertation, Dept. of Civil Engineering, The Ohio State University, Columbus, Ohio.

Argialas, D.P., Lyon, J.C., & Mintzer, O.W. (1988). Quantitative description and classification of drainage patterns, *Photogrammetric Engineering and Remote Sensing*, vol. 54, no. 4, pp. 504-509.

Colby, J.D. (1991). Topographic normalization in Rugged terrain, *Photogrammetric Engineering and Remote Sensing*, vol. 57, no.5, pp.531-37.

Dubayah, R.O., & Dozier, J. (1986). Orthographic terrain views using data derived from digital terrain models, *Photogrammetric Engineering and Remote Sensing*, vol 52, no 4, pp.509-518.

Hadipriono, F.C., Lyon, J.G., & Li W.H., T. (1990). The development of a knowledge-based expert system for analysis of drainage patterns, *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 6, pp. 905-909.

Howard, A. (1967). Drainage analysis in geologic interpretation: a summation, *American Association of Petrolium Geologists*, vol. 51, pp. 2246-2259.

Jenson, S.K (1985). Automated derivation of hydrologic basin characteristics from digital elevation model data, Proceedings of Auto-Carto 7, Washington, D.C., pp. 301-310.

Jenson, S.K., & Domingue, J.O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogrammetric Engineering and Remote Sensing*, vol. 54, no. 11, pp. 1593-1600.

Lee, J., Snyder, P.K., & Fisher, P.F. (1992). Modeling the effect of data errors on feature extraction from digital elevation models, *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 10, pp. 1461-1467.

Musavi, M.T., Shirvaikar, M.V., Ramanathan, E., & and Nekovei, A.R. (1988). A vision based method to automate map processing, *Pattern Recognition*, vol. 21, no. 4, pp. 319-326.

O'Callaghan, J.F., & Mark, D.M. (1984). The extraction of drainage networks from digital elevation data, *Computer Vision, Graphics, and Image Processing*, vol. 28, pp. 323-344.

Parvis, M. (1950). Drainage pattern significance in air photo identification of soils and bedrocks, Photogrammetric Engineering, vol. 16, pp. 387-409.

Qian, J., Ehrich, R.W., Campbell, J.B. (1990). DNESYS - An expert system for automatic extraction of drainage networks from digital elevation data, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 1, pp. 29-44.

Wiesel, J.W. (1985). Digital image processing for orthophotogeneration, *Photogrametrica*, Vol. 40, No. 2, pp. 69-76.

--- Military Specification, Digital Terrain Elevation Data (DTED) Level 1, MIL-D-89000, Defense Mapping Agency, ATTN: PRS, 8613 Lee Highway, Fairfax, VA 22031-2137, April 1986.

--- Khoros Image Processing Software, University of New Mexico, Department of Electrical and Computer Engineering, 1990.

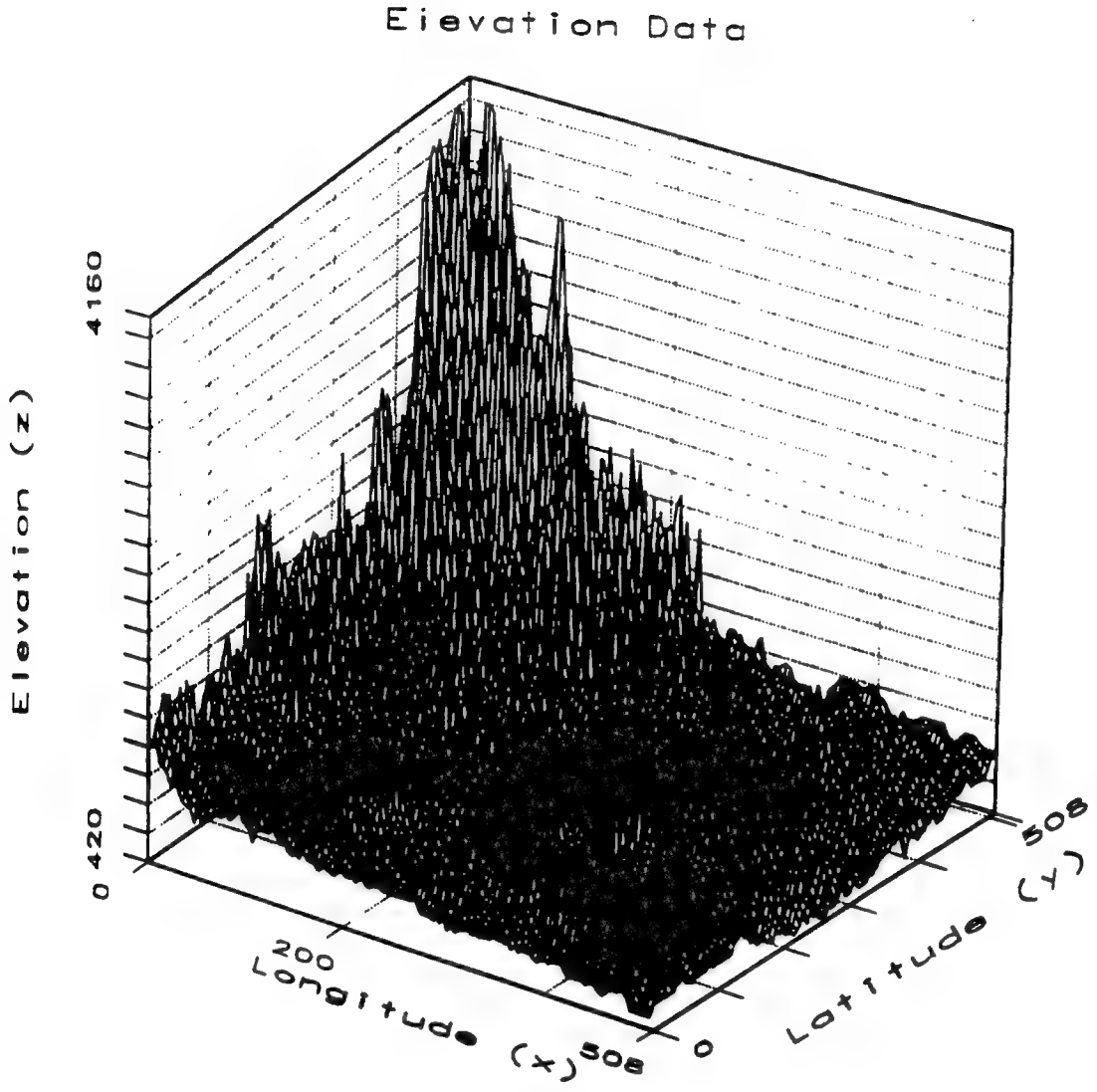


Figure 2. A three dimensional plot of a sample elevation data.

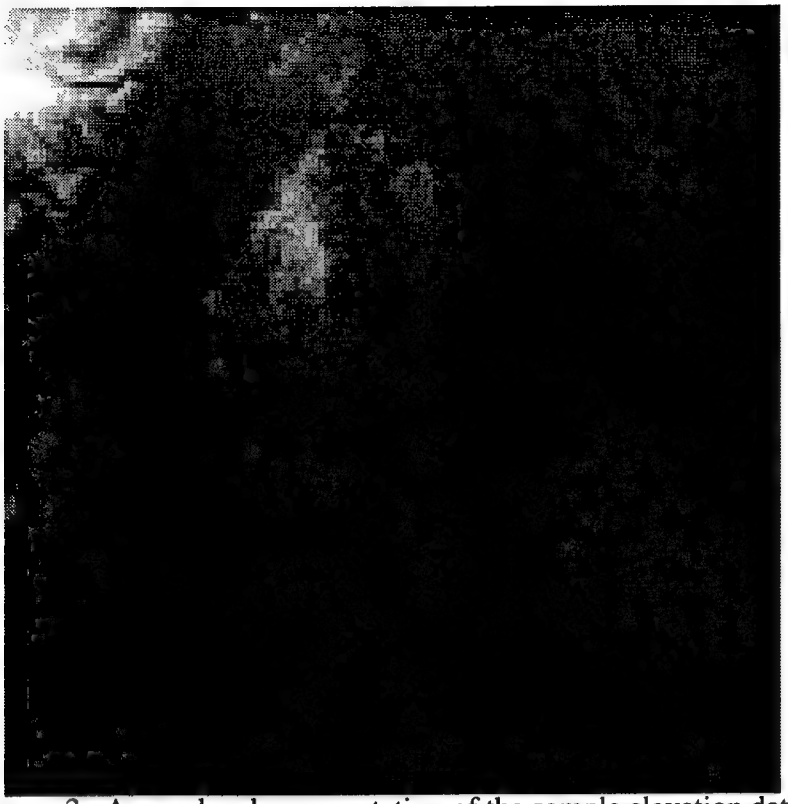


Figure 3. A gray level representation of the sample elevation data.

Figure 4. The result of a high pass filter on the sample elevation data.

Figure 5. The result of a histogram equalization on data of Figure 4.

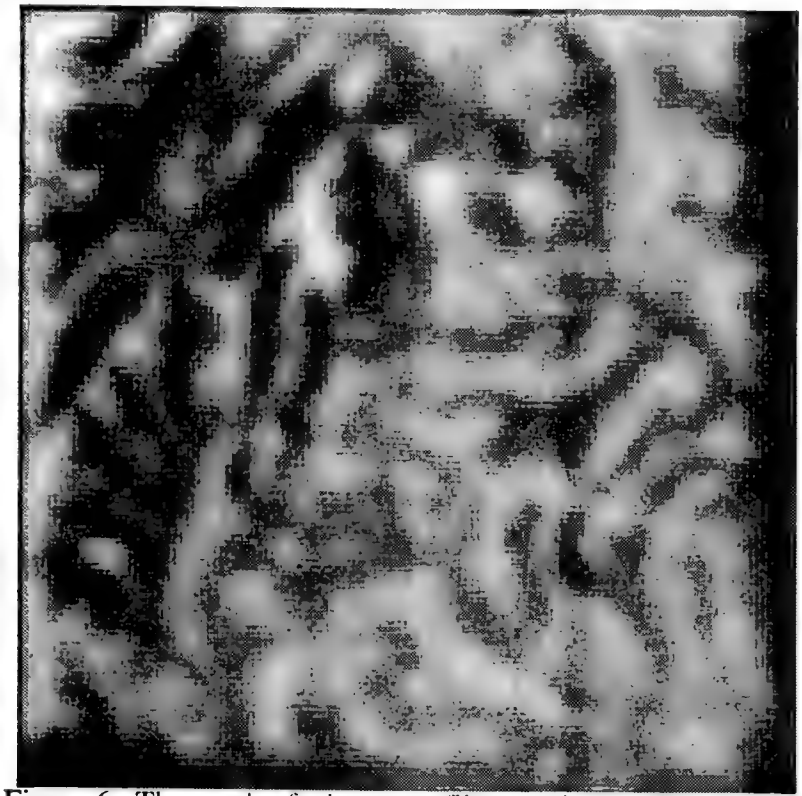


Figure 6. The result of a low pass filter on the equalized image.

Figure 7. The result of a variable threshold binarization on the data of figure 6.

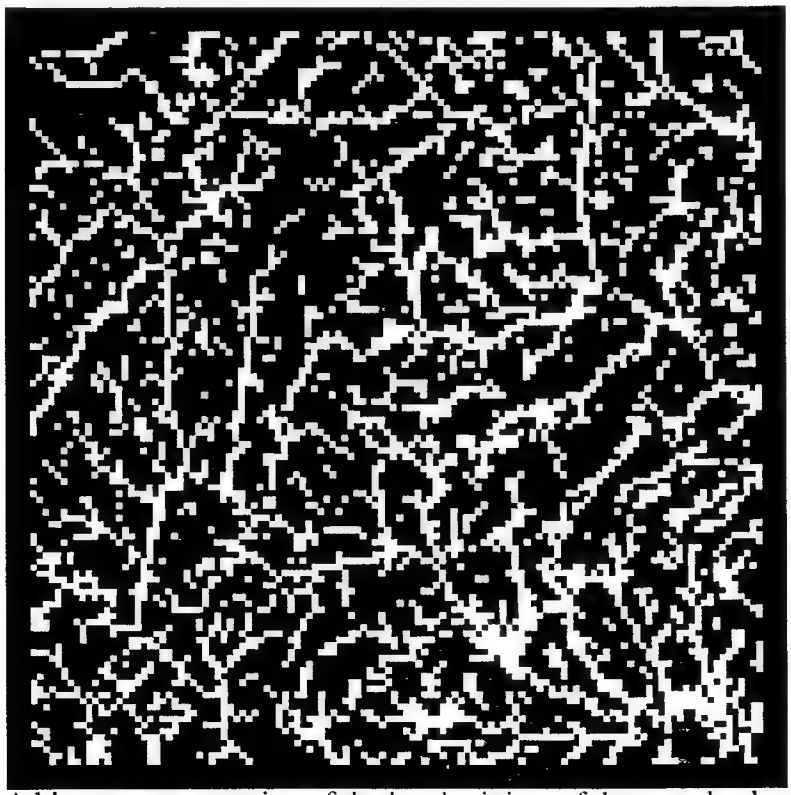


Figure 8. A binary representation of the local minima of the sample elevation data.

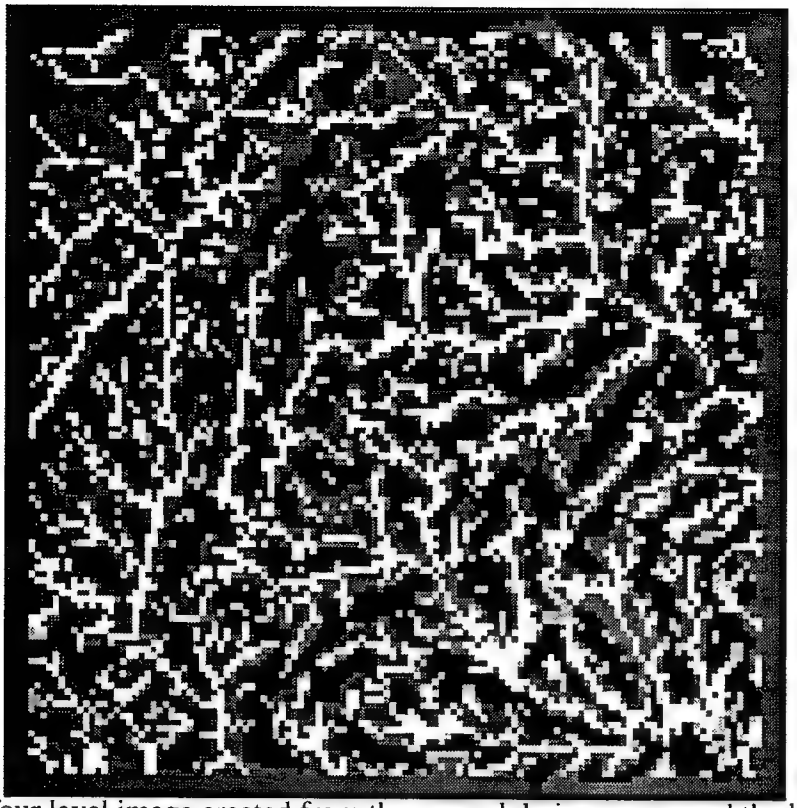


Figure 9. A four level image created from the general drainage area and the local minima.

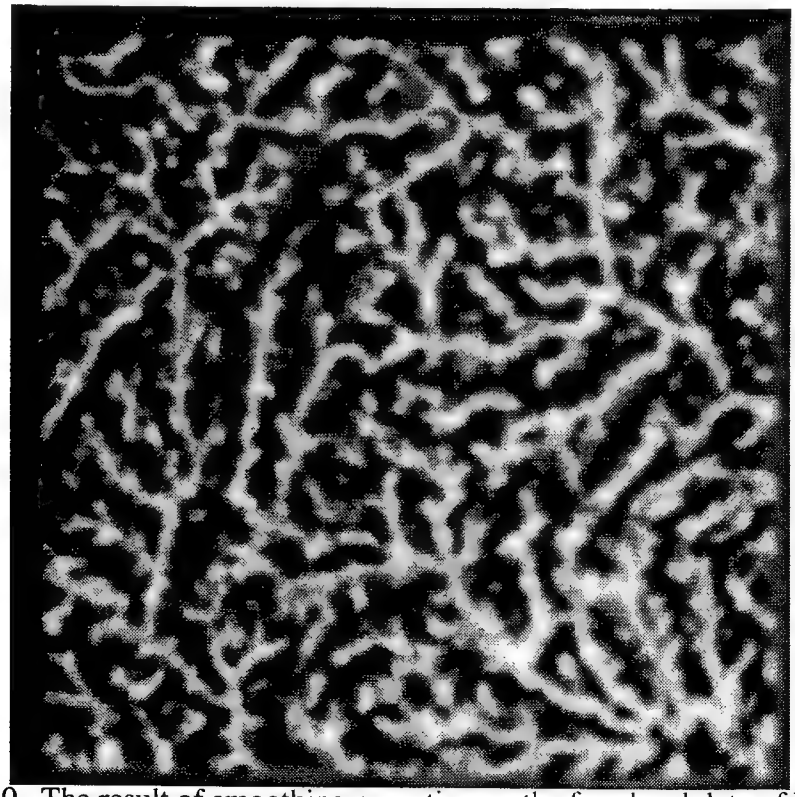


Figure 10. The result of smoothing operation on the four level data of Figure 9.

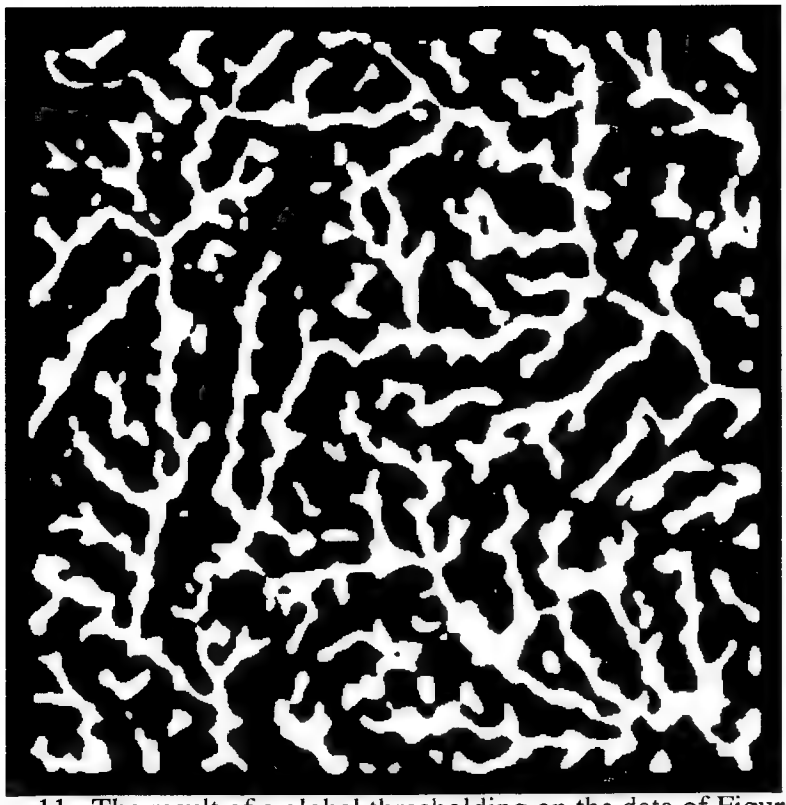


Figure 11. The result of a global thresholding on the data of Figure 10.

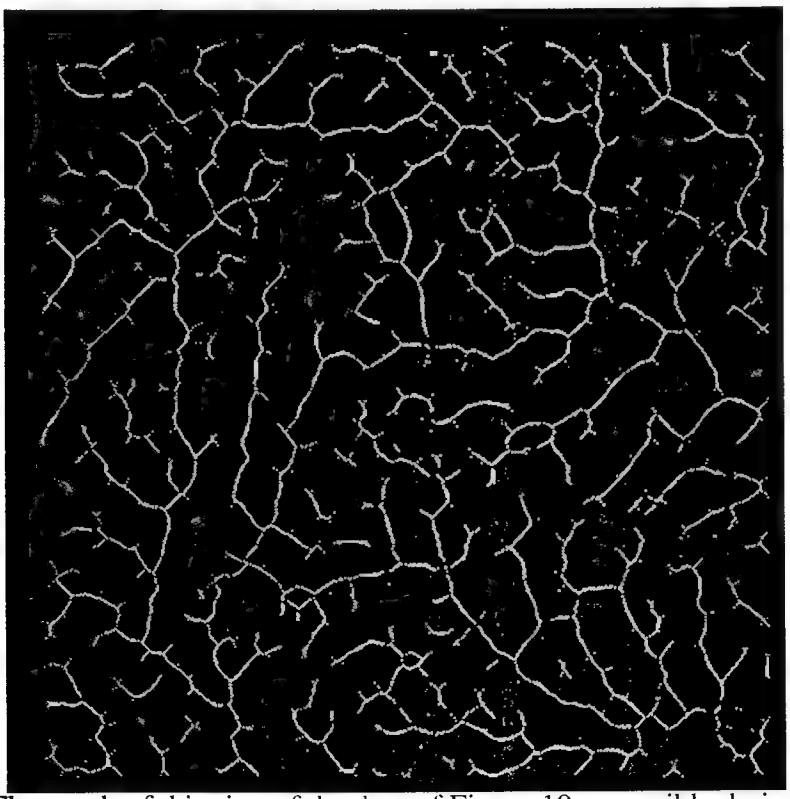


Figure 12. The result of thinning of the data of Figure 10, a possible drainage network.

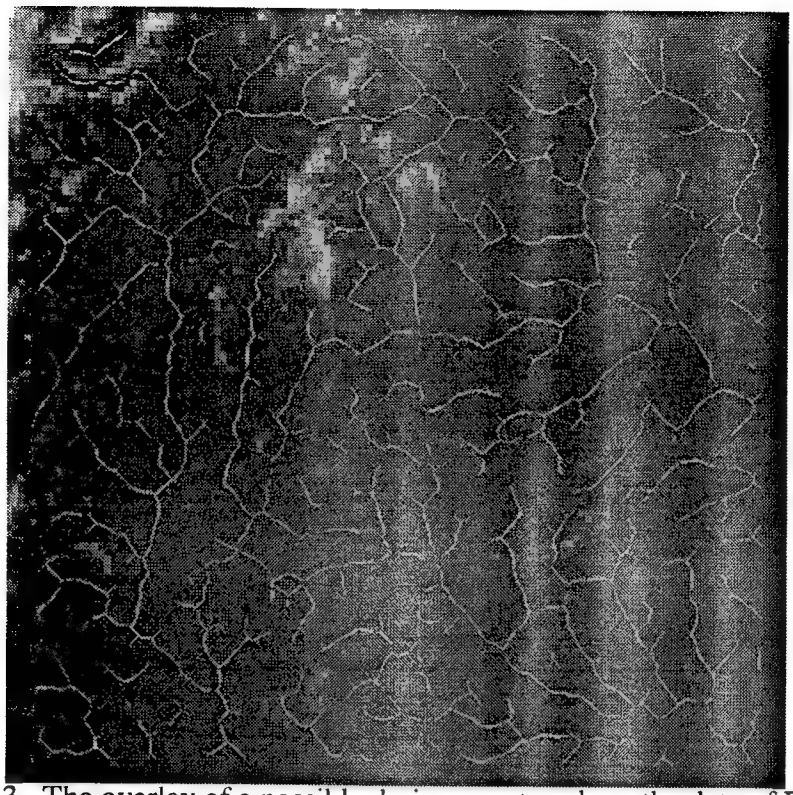


Figure 13. The overlay of a possible drainage network on the data of Figure 4.

Mutual Coupling Effect of Square Microstrip Patch Antennas on a Ferrite Substrate

Daniel K. Lee
Graduate Student
Department of Electrical Engineering

Southern Illinois University - Carbondale
Carbondale, IL 62901-6603

and

Frances J. Harackiewicz
Assistant Professor
Department of Electrical Engineering

Southern Illinois University - Carbondale
Carbondale, IL 62901-6603

Final Report for:
Graduate Summer Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

Mutual Coupling Effect of Square Microstrip Patch Antennas on a Ferrite Substrate

Daniel K. Lee
Graduate Student
Department of Electrical Engineering
Southern Illinois University - Carbondale

Frances J. Harackiewicz
Assistant Professor
Department of Electrical Engineering
Southern Illinois University - Carbondale

Abstract

This report presents experimental investigations of the mutual coupling effect between square microstrip patch antennas on a ferrite substrate. The experiment measures mutual coupling between elements mounted on thin and bulk ferrites as well as elements in different configurations, E-plane and H-plane. The mutual coupling level decreases over 10dB as the in-plane directed DC magnetic field is applied to the ferrite substrate for elements in both configurations The mutual coupling's dependence on the thickness of the substrate is also discussed.

Mutual Coupling Effect of Square Microstrip Patch Antennas on a Ferrite Substrate

Daniel K. Lee and Frances J. Harackiewicz

Introduction

The experimental measurement of mutual coupling level between L-band rectangular, nearly circular, and circular microstrip patch antennas on a dielectric substrate by a series of measurements of the S-parameters has been discussed in the literature [1,2]. The complex $S_{ij}(i,j = 1,2)$ parameters were measured as a function of element spacing for each E-plane and H-plane configuration. The actual element spacing was increased by inserting dielectric spacers between the two elements. In this report, a dielectric material is substituted with a ferrite material and the element spacing between the two elements was increased electrically by applying the DC magnetic fields. The report summarizes the mutual coupling level between two microstrip patch elements on thin and bulk ferrite substrate as a function of the bias strength. The thin-film ferrite substrate used was YIG-GGG-YIG (Yttrium Iron Garnet-Gadolinium Gallium Garnet-Yttrium Iron Garnet) with 75 micron thick YIG films on both sides of the 0.020 inch thick GGG substrate, and 1.25mm thick YIG was used for the bulk ferrite.

Experimental Setup

The overall view of the experimental setup can be seen in figure 1. S-parameter measurements were made for square patches on YIG-GGG-YIG and YIG ferrite with various bias strength. Measurements were made with two elements in either the E-plane or the H-plane configuration as shown in figure 2 and 3 respectively. Each antenna was fed with a probe and its location as well as patch size, spacing between the two elements, and the substrate dimension are also shown in figure 2 and 3. All experiments reported in this paper is based on IL-biasing. Although, the mutual coupling effect with IX-biasing was studied, it is not included in this report because IX-biasing had no effect on the mutual coupling level. Thus, it was not much of an interest. Normal biasing could not be tested because of the difficulty in producing normal directed magnetic field with the hindrance from the coax cables which protrudes directly below the patch elements.

The magnetic field was generated by using an electromagnet with 18.5 cm radius metal pole faces separated by 10.6 cm. Magnetic field strength variation between pole faces was roughly 13%. The S-parameter measurements were carried out with Hewlett Packard's HP 8510, a vector network analyzer system(VNA). The reflection coefficients seen at port 1 S11 were measured with port 2 terminated in a 50Ω load. The coupling coefficients S21 were measured with both ports connected to the VNA.

Experimental Results

S-parameter measurements with DC magnetic biasing at 300 Gauss and 500 Gauss with respect to frequency for elements on thin-film ferrite are shown in figure 4. It can be seen from the figure that the increase in the bias strength increases the resonant frequency. Also, the peak $|S_{21}|^2$ shifts together with the peak of $|S_{11}|^2$ as the biasing increases but their peak frequency does not exactly line up with one another. The measurements of peak $|S_{21}|^2$ and the resonant frequency respect to the bias strength for elements on bulk ferrite E-plane and H-plane configurations are shown in figure 5. Both graphs in figure 5 show a jump in mutual coupling followed by a gradual decrease with the increase in the bias strength. Figure 6 shows a similar result for elements on thin ferrite E-plane and H-plane configuration. On each graph, the resonant frequency with respect to the bias strength are also shown. In all cases, the resonant frequency tuning with the increase in the bias strength can be seen. One of the important factors to be considered for an accurate mutual coupling measurement between the two patch element is the contributions from the surface waves and input impedance mismatch on the mutual coupling level.

The surface wave seems to have little contribution on the mutual coupling since the mutual coupling level did not vary much (within couple of dBs) as the surface wave crossed over the radiating resonant frequency. It was pointed out as well in the literature [1] that dielectric surface waves have very small contribution to the mutual coupling level.

The contribution of mismatch to $|S_{21}|$'s can be analyzed using figures 7 and 8. Each figure has a set of three curves, $1 - |S_{11}|^2$, $1 - |S_{11}|^2 - |S_{21}|^2$, and $|S_{21}|^2$. Both $1 - |S_{11}|^2$, and $1 - |S_{11}|^2 - |S_{21}|^2$ curves which take into account of mismatch factor show little variation with the change in the bias strength. However, the $|S_{21}|^2$ curve shows a decreasing trend as the bias strength increases. Thus, the decrease in

the mutual coupling level with the increase in the bias strength does not seem to depend on the surface wave or on the mismatch factor. The typical drop of mutual coupling level was over 10dB as the DC magnetic field varied from 50 Gauss to 750 Gauss for the thin ferrite and up to 1000 Gauss for the bulk ferrite.

The Mutual coupling effect due to the substrate thickness can be seen by making a comparison between figure 5 and 6. The mutual coupling level between the two figures do not show any conclusive relationship to the substrate thickness. Thus, based on the experiments shown in this report, the mutual coupling level and the substrate thickness do not seem to have much correlation.

Conclusion

Both E-plane and H-plane configurations with thin-film and thick ferrite substrates have been tested for mutual coupling between two patch elements. It has been observed that the mutual coupling level in all cases first increases and then decreases with the increase in bias strength. The resonant frequency tuning upon biasing electrically increases the separation between the two patch antennas, thus decreasing the mutual coupling level. The results also show that the change in the mutual coupling is mainly due to the spacing between the two elements and does not heavily depend on the surface wave or on the mismatch factor.

BIBLIOGRAPHY

1] R.P. Jedlicka, M.T. Poe, and K.R. Carver, "Measured Mutual Coupling Between Microstrip Antennas," IEEE Trans. Antennas Propagat., Vol AP-29, NO.1, PP.147-149, Jan. 1981.

[2] David M. Pozar, "Input Impedance and Mutual Coupling of Rectangular Microstrip Antennas," IEEE Trans. Antennas Propagat., Vol AP-30, NO.6, PP.1191-1196, Nov. 1982.

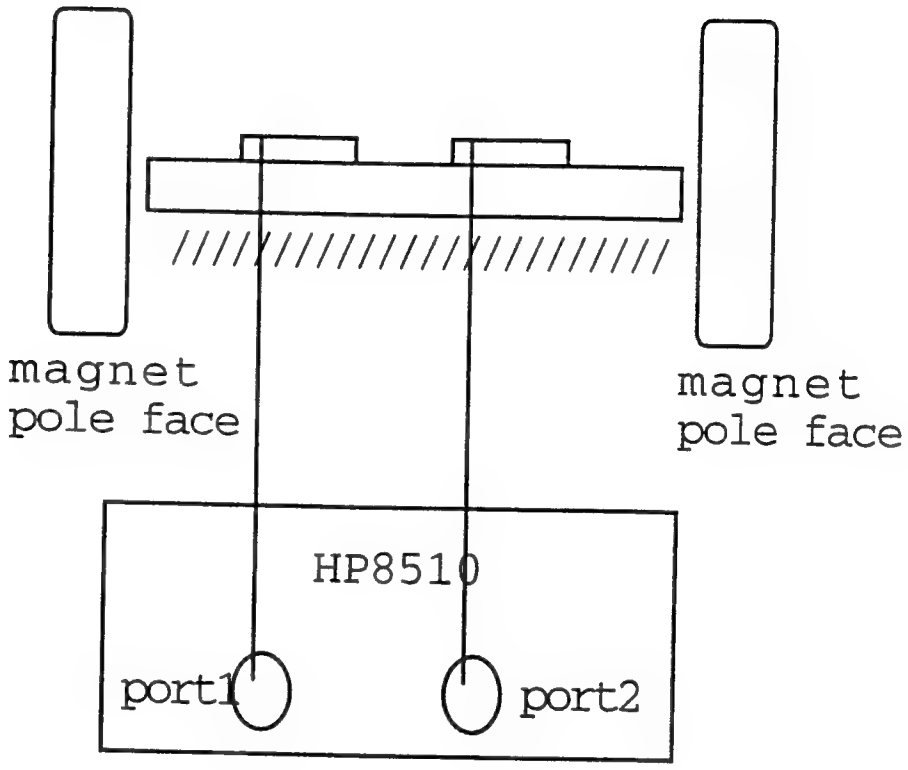


Figure 1. Mutual coupling experiment set up

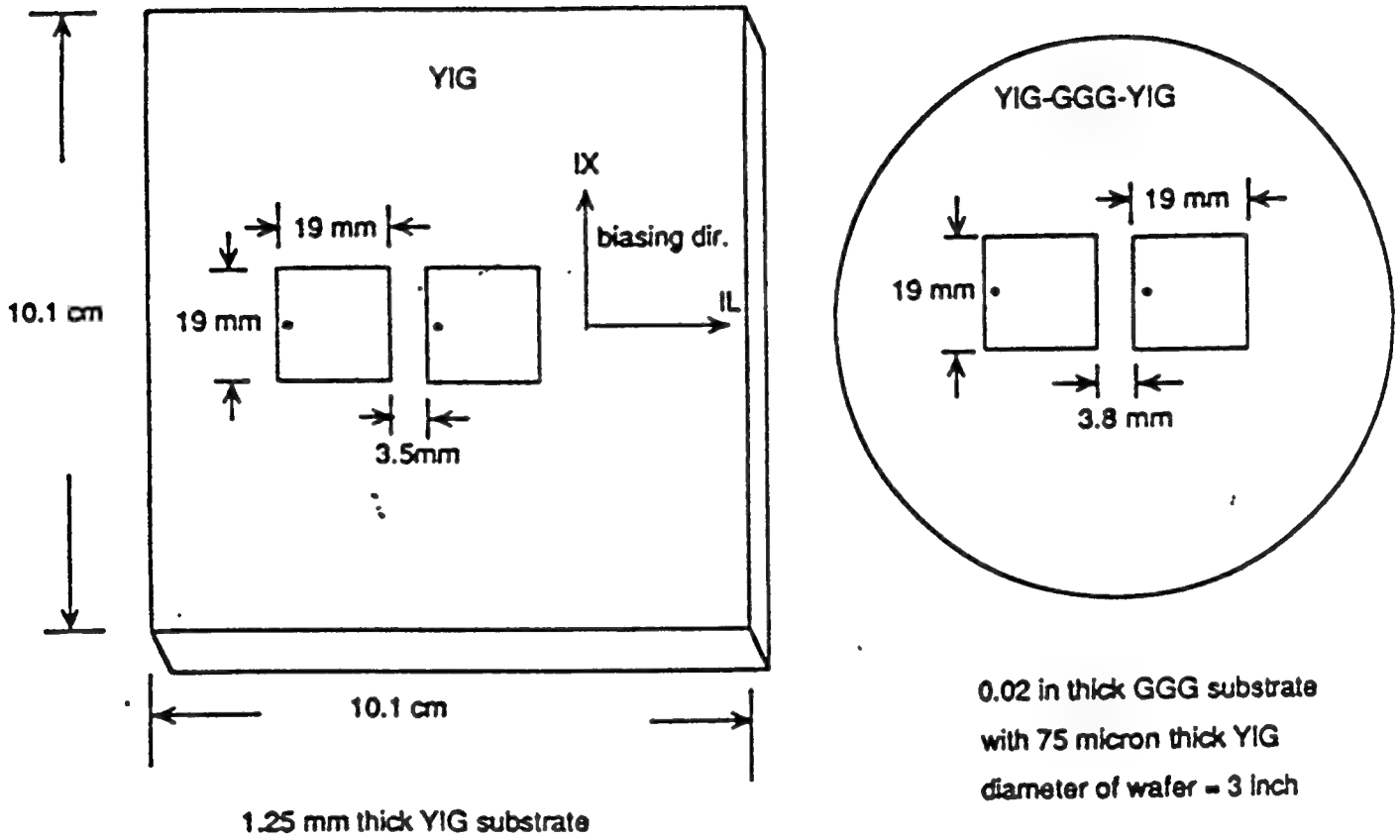


Figure 2. Patch size and spacing for elements in E-plane configuration on bulk and thin ferrite substrates

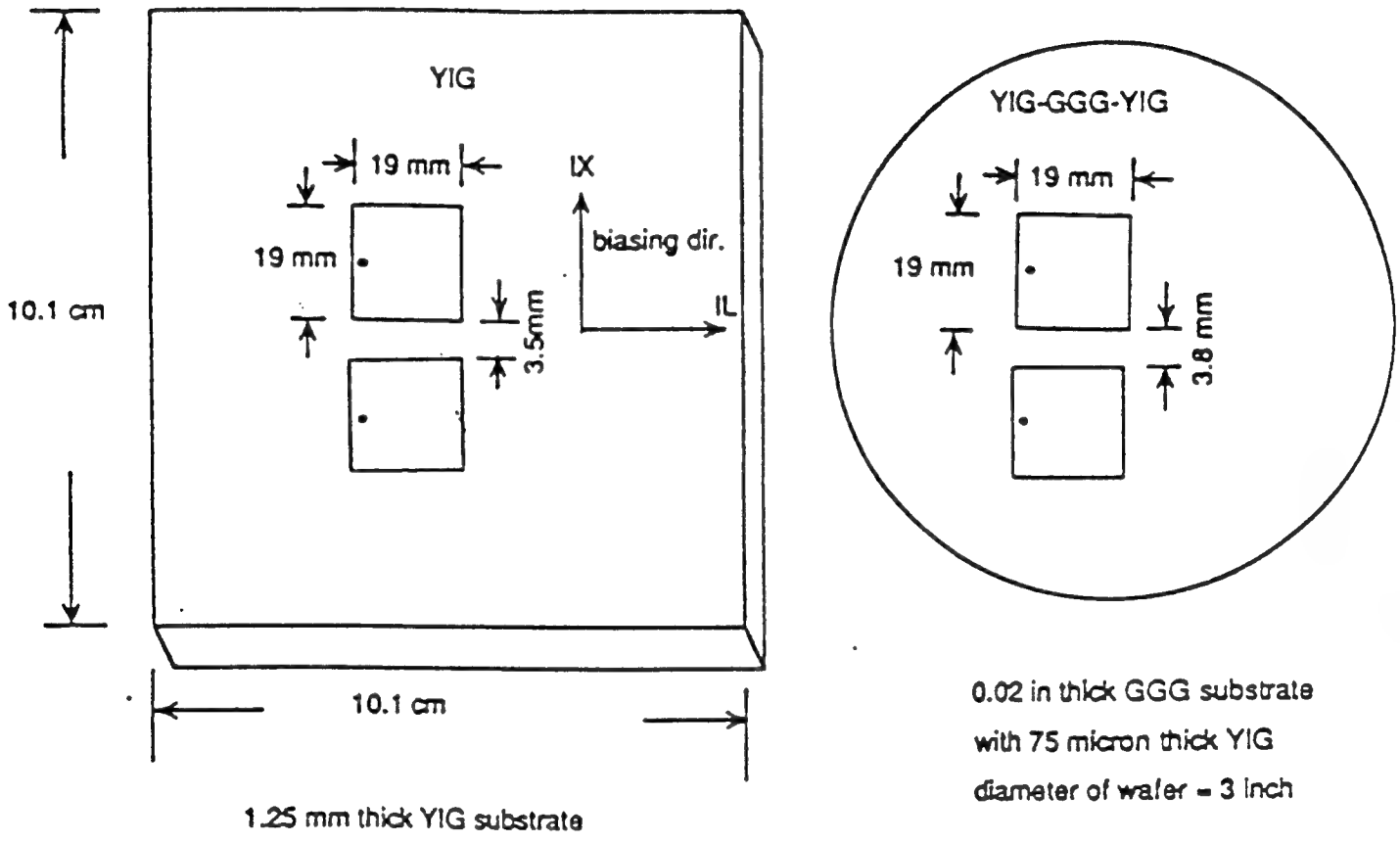


Figure 3. Patch size and spacing for elements in H-plane configuration on bulk and thin ferrite substrates

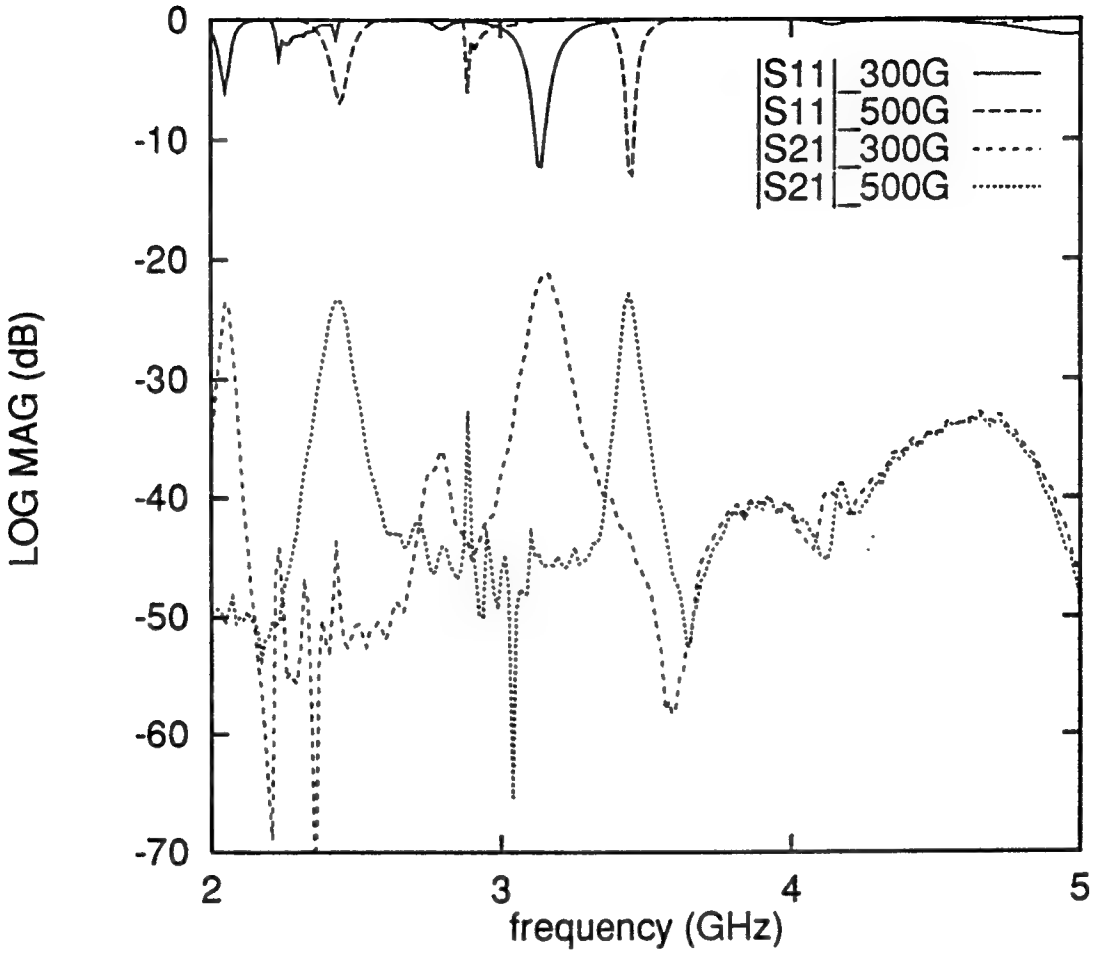
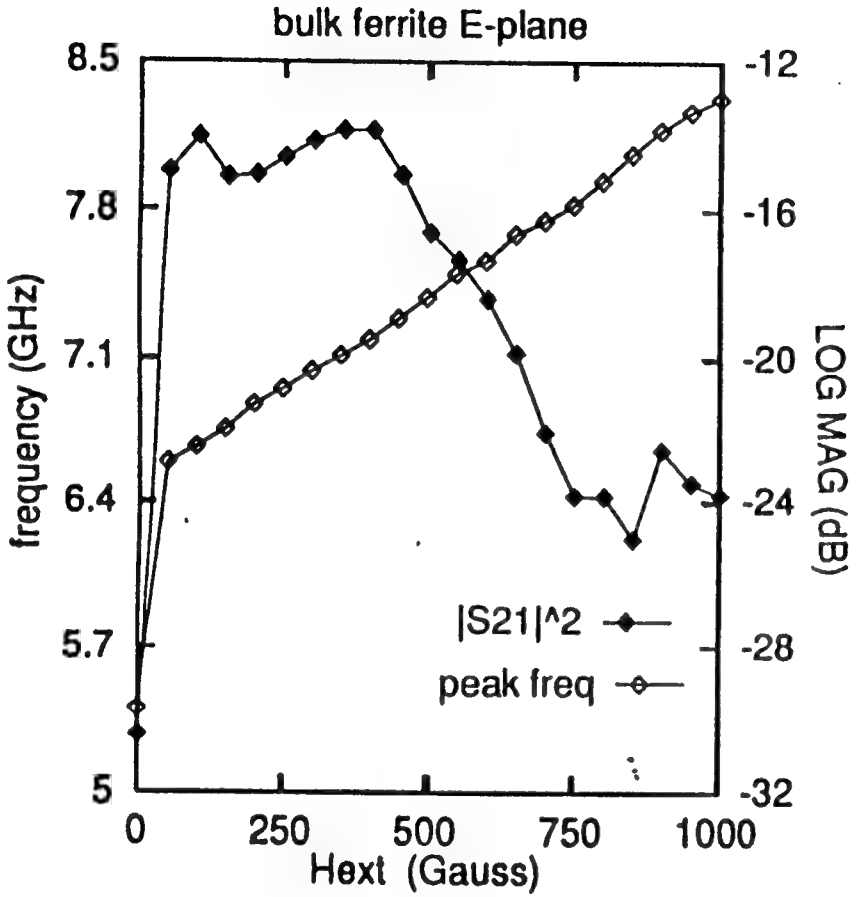


Figure 4. |S11| and |S21| vs frequency for elements on thin-film ferrite substrate with IL-biased at 300 Gauss and 500 Gauss



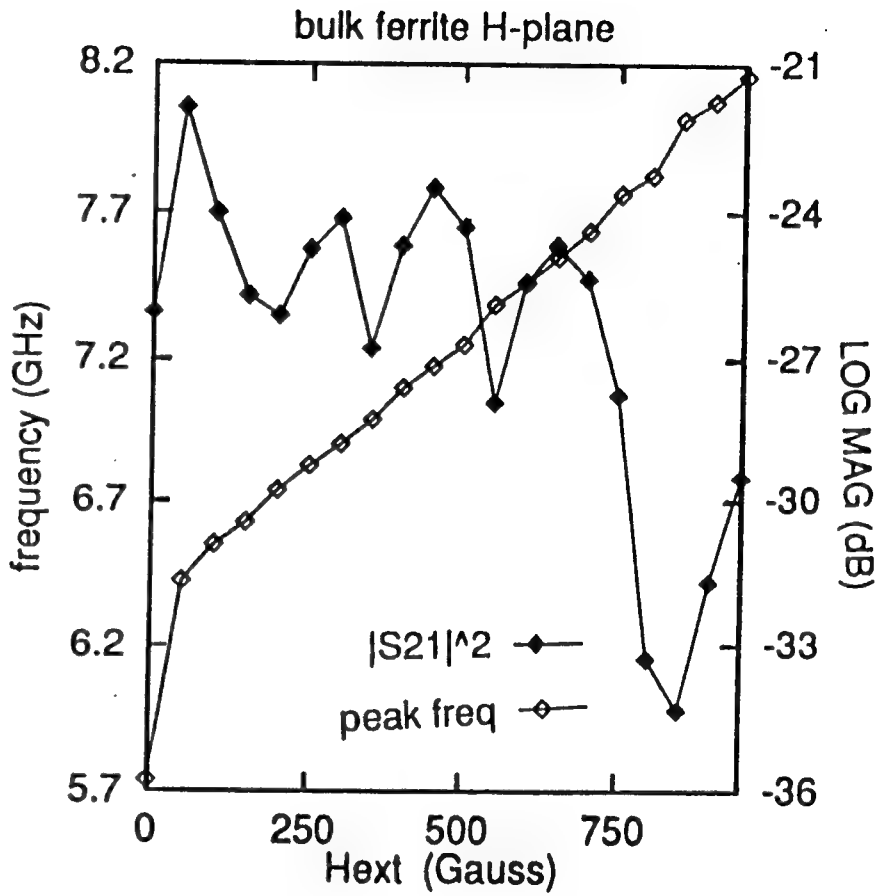


Figure 5. Resonant frequency and mutual coupling level vs.the bias strength for elements on bulk ferrite.

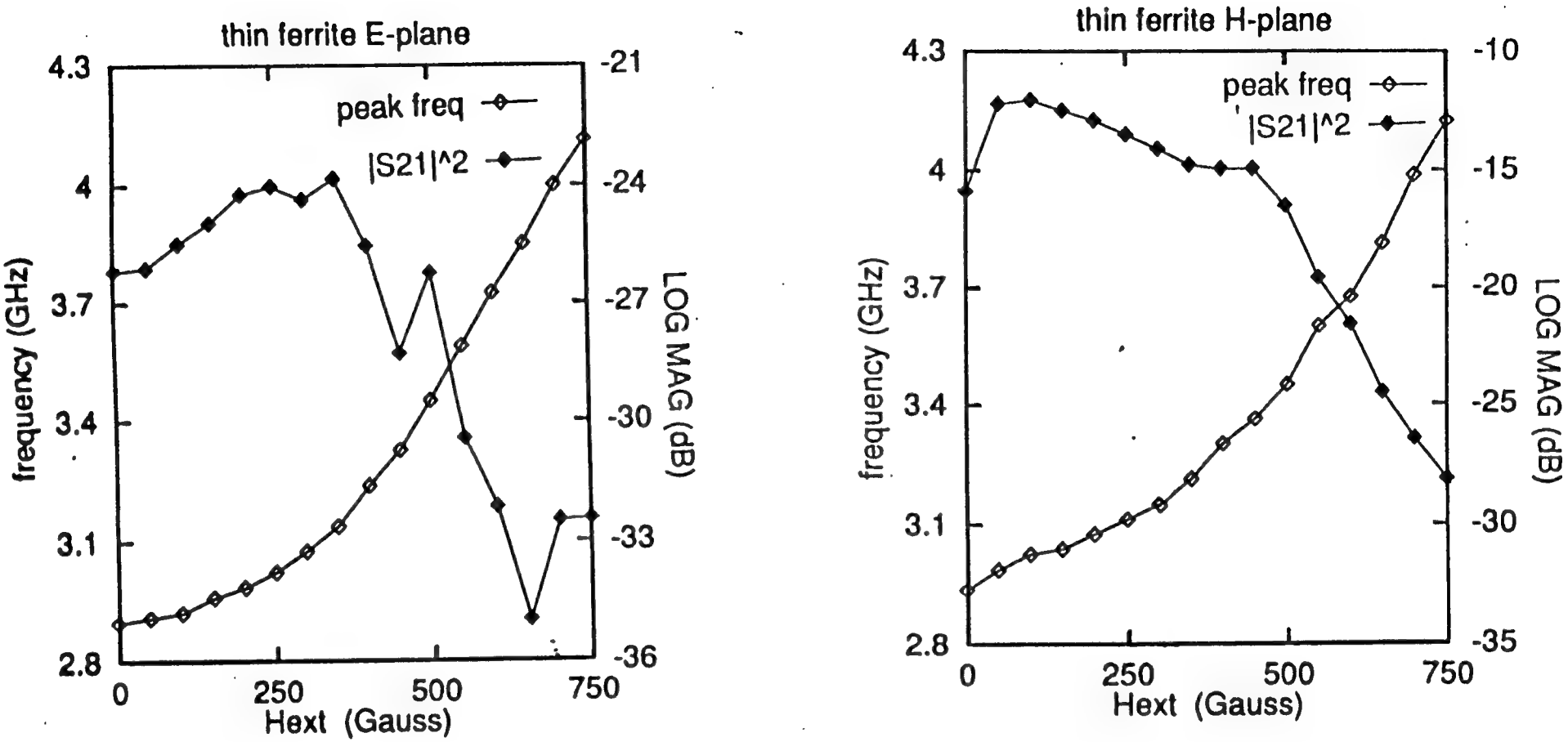


Figure 6. Resonant frequency and mutual coupling level vs. the bias strength for elements on thin ferrite.

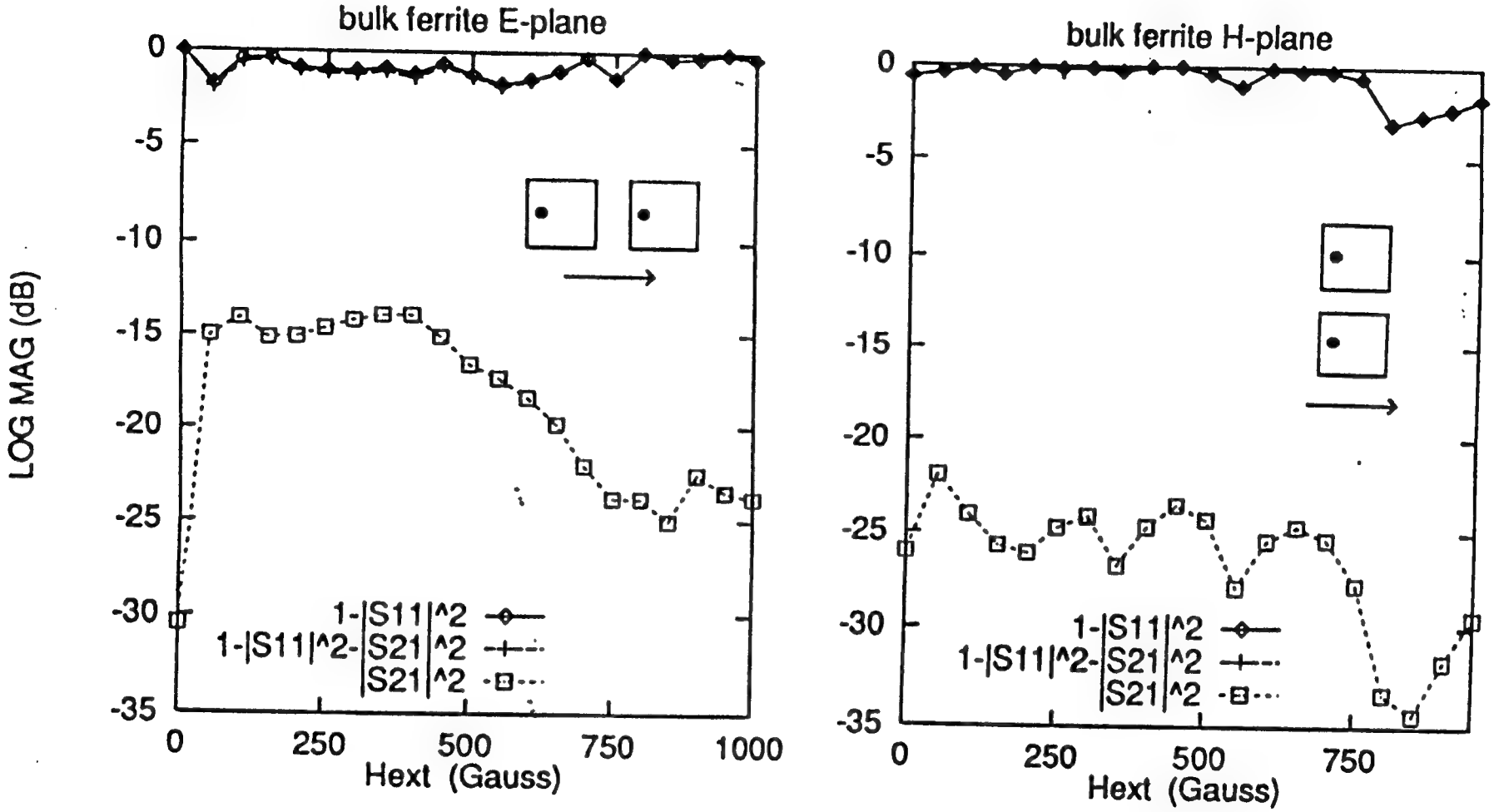


Figure 7. Mismatch factor and mutual coupling level with respect to the bias strength for elements on bulk ferrite.

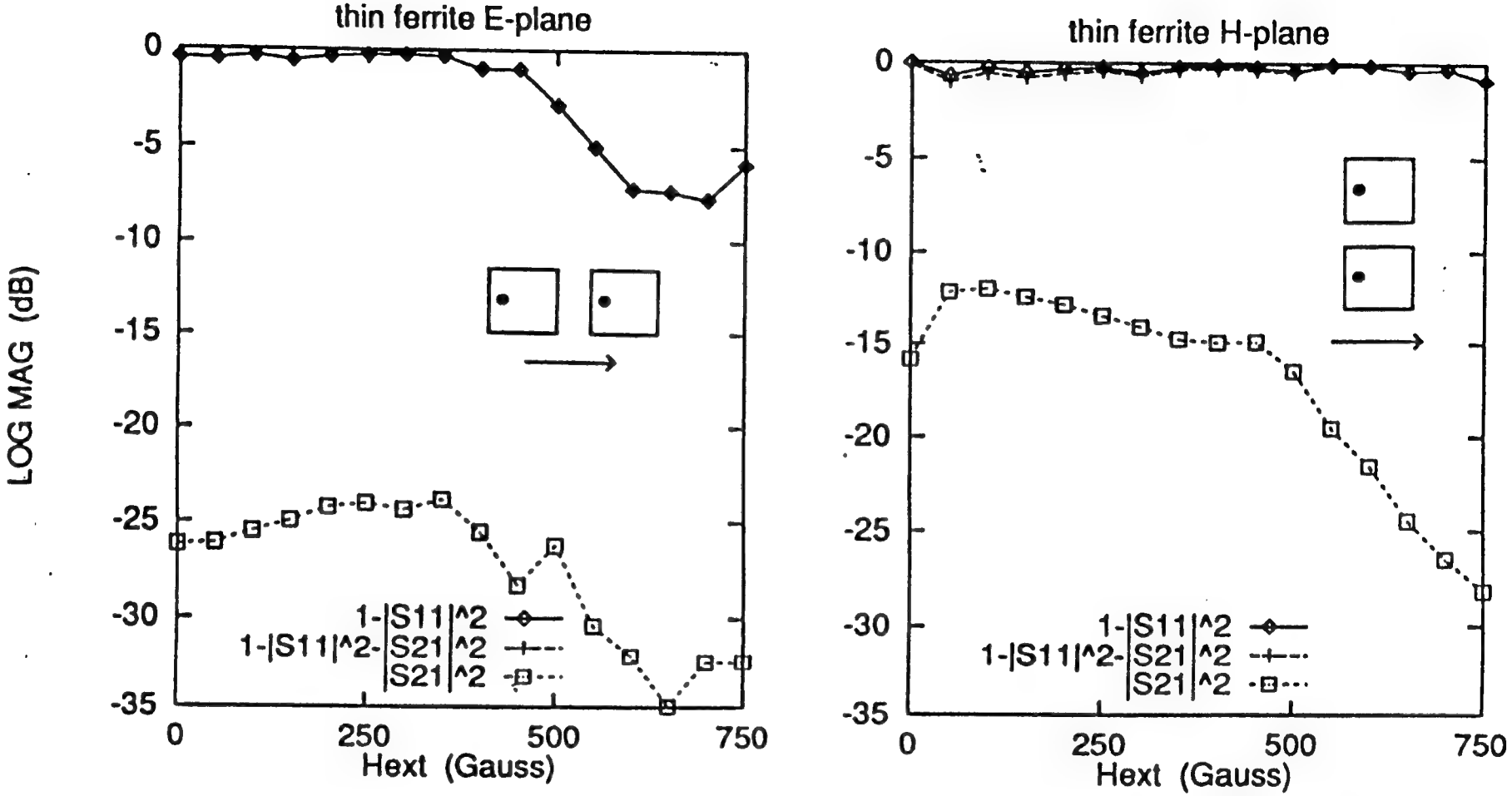


Figure 8. Mismatch factor and mutual coupling level with respect to the bias strength for elements on thin ferrite.

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Ph. D.
Professor

Slawomir J. Marcinkowski
Ph. D. Student

John Carlo Bertot
Ph. D. Student

School of Information Studies
4-293 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100

Final Report for:
Summer Faculty Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Slawomir J. Marcinkowski and John Carlo Bertot
School of Information Studies
Syracuse University

Abstract

The current global economic climate is such that a nation acquires and maintains its wealth, prosperity, and strength predominantly through trade. It is necessary, but no longer sufficient, for a nation to possess a strong military function in a global marketplace. This new emphasis on national competitiveness in trade places Rome Laboratory, as well as other federal laboratories, at an important crossroads. On the one hand, a military advantage requires continual technological superiority. On the other hand, Rome Laboratory needs to facilitate national economic development through the transfer of its technologies into the marketplace. These developments serve to highlight the importance of a proactive technology transfer process within Rome Laboratory. However, a proactive operation is difficult to put into action without forthcoming budgetary and personpower increases. This study focuses on a low-cost alternative: to use MOSAIC on the Internet and the World Wide Web to promote the transfer and commercialization of technology. An electronic system was developed allowing access on various technology transfer information and databases to the private sector, as well as Rome Laboratory and other Air Force and public sector users. This report describes these efforts, underlying reconceptualizations, design and implications of the electronic system.

TRANSFERRING TECHNOLOGY VIA THE INTERNET

Rolf T. Wigand, Slawomir J. Marcinkowski and John Carlo Bertot

Introduction

The current global economic climate is such that a nation acquires and maintains its wealth, prosperity, and strength predominantly through trade. It is necessary, but no longer sufficient, for a nation to possess a strong military function in a global marketplace. This new emphasis on national competitiveness in trade places Rome Laboratory (RL), as well as other federal laboratories, at an important crossroads. On the one hand, a military advantage requires continual technological superiority. On the other hand, RL needs to facilitate national economic development through the transfer of its technologies into the marketplace. These developments serve to highlight the importance of a proactive technology transfer operation within RL. However, a proactive operation is difficult to put into action without forthcoming budgetary and personpower increases. This study attempts to identify a low-cost alternative: to use the Internet to promote the transfer of technology.

RL is currently positioning itself to meet the challenges precipitated by the emphasis on the technology transfer process. This process, if viewed as a life cycle, minimally contains the following components:

Technology research

Technology development

Technology announcement

Technology marketing

Fulfilling technology information requests

Making the technology available for outside use

Technology utilization, and

Technology commercialization.

In addition, technology transfer occurs within political and social climates that may impose certain constraints. The overriding condition in all of these considerations is, of course, first of all the Air Force

technology transfer to the private sector, hopefully resulting in technology commercialization. The latter term incorporates such desirable events as successful technology transfer, technology utilization, profitability of the technology when marketed by a firm, hopefully resulting in the creation of jobs, adding value, contributing to the regional and national economies, and adding overall to the Gross Domestic Product.

Many of these technology transfer tasks are mandates and specified by legislation such as the Stephenson-Wydler Act, the Technology Transfer Act, and others. There are also certain limitations and constraints on the technology transfer process such as export control regulations intended to help only U. S.-owned companies.

Technology transfer deliverables could be classified as: products, knowledge, advice, know-how, and facilities use. These deliverables are not unique to RL, but can be envisioned to apply to all federal laboratories. They can be broken down into a set of specific and clearly definable technology transfer activities may include the following:

Cooperative Research and Development Agreements (CRDAs)

Educational Partnerships

Exhibits

Conferences

Presentations (by the Technology Transfer Office or scientists and engineers)

White papers

Information exchange

Inquiries (face-to-face, by phone, written)

Publications

Advertising

Public relations

Communication with company representatives

Electronic dissemination

Others.

The present focus for this study is on just one of the above activities, electronic dissemination of information about technology.

The authors reviewed the technology transfer process at RL and revisited this process by reconceptualizing the underlying assumptions pertaining to basic principles of marketing, diffusion, information retrieval and networking. The key driver of this reconceptualization was that of a user or target audience orientation. From this perspective a model was developed that provides the desired electronic linkage between RL and targeted private sector firms.

Reconceptualization

Four different approaches to a reconceptualization of the technology transfer process, especially when this process occurs electronically, were advanced. Each is described below:

A. Marketing. All marketing efforts are based on the basic premise that there is a specific audience of consumers. This is a set of specific firms within an industry or the private sector, which have needs that can be filled by firms operating within a specific market. In the this case, RL constitutes such a firm that can fill the demand for specific technologies that can be marketed and commercialized by private sector firms.

One can identify three main policies of orientation within the marketing effort: customer orientation, product orientation and profit orientation. The latter orientation is not an appropriate orientation for RL. A product orientation is predicated on the view that consumers will recognize and appreciate products for superior merit and bestow their patronage on firms--in this case the product is research and development. In the present case though, it must be realized that, in general, this product is not designed by RL having private sector firms in mind foremost. Rather, the key customer is the Air Force and the product and the product orientation are determined by the Air Force mission. An exclusive product orientation effort is therefore not wise for RL, but the product certainly determines a potential market that can be tapped within the private sector. It is therefore apparent that an appropriate focus for RL in terms of marketing orientation is a customer orientation.

In this setting, a customer orientation denotes (a) an attitude and a pattern of conduct, as well as (b) the extent to which RL tries to determine what its customers want and then gives them what they want. Granted, it is often difficult and complex to discover what the customer wants. For example, in such efforts one recognizes that customer preferences vary widely and that customers want something only once they are given concrete choices, i. e. they are incapable of conceiving possible new products. Potential customers often also request features that are incompatible with product capabilities.

The basic challenge faced by RL is to identify those needs and provide a linkage between RL and the customer, i. e. a firm interested in a specific RL technology. To maximize this linkage and relationship to customers, RL needs to form hypotheses and understanding about these present and future potential customers. Included should be questions such as:

What kinds of things affect customer behavior?

Through which channels (advertising, face-to-face contacts, publications, etc.) can customers be reached by RL?

What is the degree or strength of need or desire for the product?

What are the appropriate appeals (or arguments) to which customers are most responsive?

What is the customer's responsiveness to different types of sales devices, i. e. their ability to be influenced by technology transfer discussions by the Technology Transfer staff, engineers and scientists?

After these questions have been answered, the marketing dimension entails five general activities:

1. Identifying and selecting the type of customer that RL chooses to cultivate and learning that set of firms' needs and desires.

2. Designing products, know-how and services, to include facilities, that RL can transfer in conformity with customer desires.

3. Persuading customers to acquire and adopt RL's products, know-how and services.

4. Displaying, moving and to some extent storing products, know-how and services after they have been developed at RL.

5. Identifying dual use technologies and applications.

In designing its products, acquiring and developing know-how, and deciding how much and what specific services to offer, RL will unquestionably benefit from having a clear picture of its *target customer*. The present project attempts to provide a linkage, an *electronic marketing channel*, between these target customers and RL.

B. Diffusion. Diffusion is another conceptualization that guided the study. Diffusion is the social process by which an innovation is communicated through certain channels over time among members of a social system (Rogers, 1983). This process clearly resembles the task at hand for RL in the technology transfer process: i. e. a technology or an innovation within RL, needs to be communicated to a set of firms (members of a social system) within a particular industry. The communication channel chosen for the present study is that of electronic messaging via the Internet. The authors view such electronic diffusion means as timely, cost-effective, quick and targeting specific means to reach the target customer described above.

Many other dimensions could be addressed here that determine the speed with which such diffusion occurs. Obviously the quicker the target audience can be reached and the significance of the technology to be transferred can be communicated, the more effective the diffusion process. The rate of diffusion or adoption of that technology is, however, determined by the characteristics of the innovation (e. g., the relative advantage, compatibility, complexity, trialability and observability).

The communication channel chosen, electronic messaging via the Internet, determines in part the success of making a successful linkage with the target customer. General diffusion principles posit that, ideally, successful communication occurs in a face-to-face setting with the target audience or customer. This, however, is often impracticable or too expensive. Substitutes to face-to-face settings are chosen and the mass media come into play. The use of mass media in such efforts, as in advertising, is often a vague undertaking, as one cannot be sure that one's message truly reaches the intended target. Often a compromise between the two extremes is desirable, e. g., the use of mass media, followed up by face-to-face meetings with smaller groups, etc. Electronic messaging via the Internet comes close to this ideal,

i. e. using a cost-effective mass medium and still reaching specific individuals. Moreover, the Internet's use makes it possible to be interactive with specific individuals within the target group and thus allowing *almost* face-to-face interaction. This potential for interactivity certainly makes the medium highly attractive as requests, needs, etc. certainly can be customized. Moreover, such interactivity makes possible the often missing *feedback* in the technology transfer process. Such feedback allows us to shape and additionally customize the very next step in the diffusion and communication process. Such customization is almost impossible when viewing the diffusion process via advertising as a communication channel.

C. Information Retrieval. In an electronic system, such as using the Internet, it is obvious that the users can readily retrieve information once the information is stored on the system. There are many data bases available at RL that could be accessed through the Internet, and many of those pertain to technology transfer and related issues. The authors made every effort to identify such data bases and make them available in a format conducive to outside users, hopefully firms that might acquire a particular product, know-how or service. The design of such information, data bases, etc. often determines their successful use. In this sense, a customer or *end-user* orientation was applied in the design of the system. Unless such design is inviting, encouraging, timely, informative and user-friendly, the success of the system is highly questionable (Taylor, 1986).

D. Networking. The topic of networking as a conceptualization embeds all three of the above topics, i. e. marketing, diffusion and information retrieval. Without these three, networking would be impossible. Networking in this sense goes beyond the traditional means of reaching a target customer. George Kozmetsky, director of the IC2 Institute, Austin, Texas, stated at a recent conference on commercialization of technology from federal laboratories that, "The secret to successful commercialization of technology developed in the federal labs is networking, networking, networking (Kozmetsky, 1993)." This importance was already demonstrated in an empirical NSF-funded study by Wigand focusing on technology transfer issues within the microelectronic industry and the role of industry, government and universities (Wigand and Frankwick, 1989).

Networking denotes interaction, feedback, customizing and creating a dialogue to a degree that otherwise would not exist. In that sense and in this particular case, RL may conceive of transferring technology, at

least in the beginning stages, as a partially cooperative means with potential target customers within industry. Networking would suggest the creation of Listservs, electronic bulletin boards (discussion groups), direct electronic inquiries that might be shared with the rest of likewise interested individuals/firms around a particular product, know-how or technology. Such individuals within firms appear to be quite interested in such dialogue as a technology emerges, although it is understandable that such dialogue would be reduced or cease entirely once the technology becomes highly marketable or commercializable due to competitive reasons. But even if RL could speed up the development and diffusion of a technology to such a point and if interaction would then discontinue, RL still would have provided a most valuable service to the development of that technology and industry. Typically, such networks do not come about by themselves. Such interaction must be fostered, encouraged , managed and nurtured by knowledgeable professionals within RL, perhaps among the Technology Transfer staff or the Directorate's Technology Transfer Focal Points. Networking, in this sense, is conceived as the highest level of technology transfer along the electronic medium of using the Internet. Lastly, networking as envisioned here, is characterized as rating high in terms of customer interaction, as well as high in terms of information content, implying that RL has the means to get closer to the customer. Moreover, it is interactive, multifaceted, customized, and on demand with time delay and time zone shifts playing merely a minor role.

Methodology

Various methods were employed to investigate and research the task at hand. The authors conducted interviews with numerous individuals within RL and with non-RL employees. Interviews were conducted with the members of the RL Technology Transfer Office staff, including RL elements at Hanscomb AFB, as well as several engineers and scientists. One engineer was interviewed in greater depth, as his ACT technology was highlighted for demonstration purposes on the Internet, i. e. the World Wide Web (WWW) at RL. This particular technology was demonstrated to show the various capabilities and cross-linkages possible using MOSAIC as the software to navigate on the WWW. In addition, several branch chiefs were interviewed as well as several area company representatives (Kaman, TRW and Booze, Allen & Hamilton). Moreover, in-depth interviews were conducted with the officers of the Photonics Development Corporation

in Rome, NY, as well as members of the RL Library. The authors interviewed appropriate, selected private sector firms in terms of their willingness, availability and readiness to enter the Internet and WWW world. The effects of the interviews were surprising to some extent as we saw a *snowball effect* emerging. That is the discussion of our interests in the electronic transfer process generated numerous additional suggestions, existing and potential services, as well as data bases. All in all, a surprising synergy was created between the research team and RL members engaged in research and technology transfer.

The Internet and the World Wide Web

The Internet is a very fast growing organization over the last several years. No one knows for sure the exact number of users, as the Internet itself is an organization of loosely coupled networks. An August 4, 1994 report of the Internet Society, the organizing body representing the Internet, reports dramatic growth in 1994. The latest Internet Society measurements reveal that there are worldwide 3.2 million reachable machines. This is an increase of 81 percent for the past year and represents an even steeper than normal increase over the past six months. Indeed, one million new hosts were added during the first six months of 1994. Of the world-wide total, the U. S. connections account for 63 percent (Internet Society, 1994). The fastest currently growing segment on the Internet are businesses, again demonstrating the potential to reach the private sector via the Internet. Although it is difficult to specify a threshold at which a critical mass of business Internet users has been reached, it is possible today to reach most major firms via the Internet. Small and medium sized firms are representative of the vast majority of firms signing on the Internet (Cf., e. g., with "Going *I-way?*," 1994; McBride, 1994).

An important caveat needs to be expressed here: Just getting a presence on the Internet is not enough. Just putting one's server on the Internet is analogous to hanging a shingle on a rural road. If one is selling something and has something important to offer that is intended to reach a large audience, it is important to have not just a shingle, but a major advertising board on the *main drag*, i. e. a major highway.

The Internet's rapid proliferation suggests that organizations in the public and private sectors want to take advantage of the Internet's interactive or two-way capabilities. In turn, this suggests the establishment and exploitation of visible and invisible social and physical networks linking and bridging public and private

sector organizations.

The appropriate use of the Internet may result in such desirable outcomes as: obtaining support from vendors, participating in joint development, collaborative research projects, providing customer support, marketing and product/service/know-how distribution. Due to space limitations it is impossible here to highlight various features of a value chain derived from appropriate Internet use. This value chain can be segmented into internal operations, inputs from suppliers, and customer relations. The Internet in conjunction with other appropriate technology and software, such as the World Wide Web and MOSAIC, lends itself ideally to provide an effective, quick and responsive interactive linkage between RL and its customers, i. e. various private sector firms (Cf., e. g., with Blattberg et al., 1994; Dallaire, 1993). Having a home page on the World Wide Web is not sufficient; the information content must reflect the potential user's information need. The authors have incorporated into the technology transfer home page the information that is most sought after by potential adopters of RL's technologies. The nature of this linkage is ideal in terms of cost-effectiveness and time-based considerations. Moreover, it provides the potential for feedback, the essential component in any effective communication effort.

Data Bases and Electronic Features Available Via the Internet and WWW

During the summer months of 1994 the authors designed, built and applied various data bases and services of interest in the technology transfer process. As mentioned initially, our efforts were driven by an overall customer/user orientation. The designed system uses MOSAIC on the WWW, is fully accessible to outside, i. e. non-RL, users, as well as RL users and features the following:

What's Happening in Tech Transfer at Rome Lab?

Technology Transfer at Rome Laboratory

Hot Products

Technology Transfer Inquiries

A Model CRDA

CRDA Data Base

Educational Partnerships

Fact Sheets of Directorates

Facilities

Skills at Rome Laboratory Database

Rome Laboratory Staff Directory

The Reliability Engineer's Tool Kit

Patents Granted at Rome Laboratory

The management, use and ownership of these databases poses an intriguing question. It is very important that data bases are kept up-to- date and various current features such as *Hot Products* and *What's Happening in Tech Transfer at Rome Laboratory* feature products and information that truly is current. Otherwise, the system designed will be of little use to the outside, and RL users will recognize quickly that it is not a timely device to find out information about RL and its technologies.

The authors suggest that the ownership of this system resides with the Rome Laboratory Technology Transfer Office. Access to various features should be categorized such that it is available to all users (including the private sector) or only the Rome Laboratory Technology Transfer staff or only the RL Technology Transfer staff and all RL scientists and engineers. We recommend that the gatekeepers of information as well as the implementors of information into the system are the various Technology Transfer Focal Points within each directorate.

Implications

The system designed clearly demonstrates that the technology transfer capabilities can be achieved via electronic means. This elegant design derives value for Rome Laboratory through the timely and appropriate alignment of information technology already in place (personal computers, high-speed networks, software such as MOSAIC, and access to the Internet and WWW), Rome Laboratory's goals and strategies with regard to technology transfer and its larger role within the Armed Forces and society, as well as the appropriate alignment with existing *business processes*, especially with regard to technology transfer. Through the appropriate alignment and organizational fit, it is possible to generate added value for Rome Laboratory. It is of considerable importance, though, to realize that this designed process

occurring electronically cannot be arrested by peripheral traditional processes that might be paper-based, and still assume that the value chain continues delivering the same value. Appropriate electronic alignment must occur in such peripheral processes as well.

One must realize also that implementing this effort, even though highly cost-effective when compared to more traditional means, e.g., advertising, does not come cheaply. The system as conceived must be taken on by an individual or a team of individuals knowledgeable of the technology transfer process, RL, private sector operating practices, and probably some technology background and competence would be helpful as well. The task at hand is not just the mounting of the system on the WWW, but the effort must be nurtured, customized and *mothered*. Especially, the involvement of key individuals in private sector firms with an interest in technology transfer in such electronic organizational forms as Bulletin Boards, Listservs, etc. requires sensitivity, understanding and a sense of managing groups from a group dynamics perspective. This is time-consuming and requires patience in real life, but especially when done electronically. Numerous examples, however, abound demonstrating success in such efforts.

References

Blattberg, Robert C., Glazer, Rashi, and Little, John D. C. (Eds.). *The Marketing Information Revolution*. Boston, MA: Harvard Business School Press, 1994.

Dallaire, Rene M. Data-Based Marketing for Competitive Advantage. *Data Resource Management*, Spring, 1993, pp. 46-50.

Going *I-way?* [Cover Story]. *PC Computing*, September, 1994, pp. 120-169.

Internet Society Press Release, August 4, 1994.

Kozmetsky, George. "Commercialization of Technology" (Keynote Address). Presented at the Conference on Commercialization of Technology within Federal Laboratories, Santa Fe, New Mexico, March, 1993.

McBride, James. Marketing MOSAIC: The War Has Begun. *Internet World*, October, 1994, pp. 40-43.

Rogers, Everett M. *Diffusion of Innovations* (Third Edition). New York: Free Press, 1983.

Taylor, Robert S. *Value-Added Processes in Information Systems*. Norwood, NJ: Ablex, 1986.

Wigand, Rolf T. and Frankwick, Gary L. Inter-Organizational Communication and Technology Transfer: Industry-Government-University Linkages. *International Journal of Technology Management*, 1989, 4(1), pp. 63-76.

Integration Of Optoelectronic Devices With Microwave Compatible Diamond Heat Sinks

Sean S. O'Keefe
Graduate Research Assistant
Department of Electrical Engineering

Cornell University
426 Phillips Hall
Ithaca, NY 14853-5401

Final Report for:
Summer Graduate Student Research Program
Rome Laboratory, Rome, N.Y.

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory, Rome, NY

August 1994

Integration Of Optoelectronic Devices With Microwave Compatible Diamond Heat Sinks

Sean S. O'Keefe
Graduate Research Assistant
Department of Electrical Engineering
Cornell University

Abstract

Thermal stability of an optoelectronic device, a vertical cavity surface emitting laser, is addressed by flip-chip mounting it onto a chemical vapor deposited diamond heat sink. This project used previously designed heat sink patterns to investigate their actual fabrication. Heat sinks fabricated in this study and previously fabricated lasers were then used to experimentally determine the procedure for flip-chip mounting the devices. The result of the study is a series of techniques for fabricating heat sinks and a reliable and repeatable process for flip-chip bonding the laser die onto the heat sink using indium solder.

Integration Of Optoelectronic Devices With Microwave Compatible Diamond Heat Sinks

Sean S. O'Keefe

I. Introduction

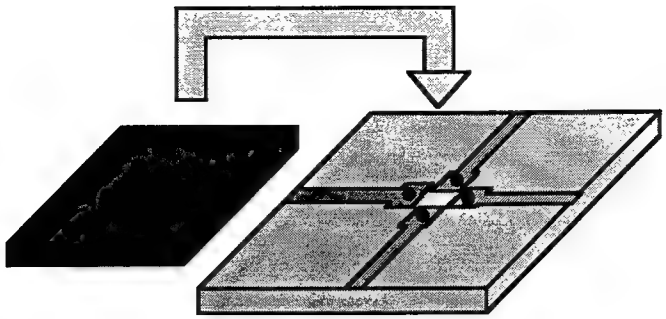
Modern computers and signal processing applications require high speed interconnects as well as high speed chips. This project investigated one aspect of a high speed optoelectronic interconnect scheme for multi-chip module (MCM) applications. Within a single multichip module, communications can be accomplished at gigahertz frequencies using stripline waveguides, but when the application is complex enough to warrant a package consisting of several stacked MCM layers, level-to-level communication becomes a serious technological problem. This project is part of a larger effort underway at the Rome Laboratory Digital Optical Processing Branch (RL/OCPB) which addresses that problem by using a through-level optical interconnect scheme. This scheme uses arrays of lasers and photodetectors on the MCM levels to send and receive information. Packaging of the transmitting elements, arrays of vertical cavity surface emitting lasers (VCSELs), was the focus of this summer project. Because of the nature of the interconnect project, the VCSEL die was to be packaged using flip-chip technology onto diamond heat sinks. Processing chemical vapor deposited diamond (CVDD) wafers into microwave-compatible heat sinks, flip-chip bonding VCSEL die onto the heat sinks, and the initial retrofitting of an evaporator to use for flip-chip solder evaporation were accomplished during the tenure of the project.

II. Discussion

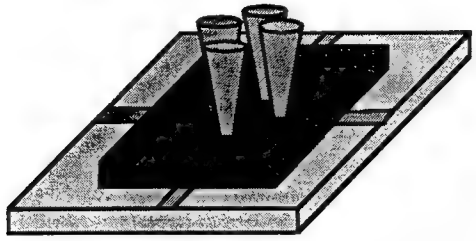
The use of optoelectronic devices in modern circuitry has its own set of special problems because practically all optoelectronic devices have very temperature-sensitive properties such as threshold current, output power, and signal-to-noise ratio. In this project, maintaining the thermal stability of one

particular device, the vertical cavity surface emitting laser, is investigated. A VCSEL is a semiconductor laser that uses high reflectivity dielectric mirrors (alternating layers of either different semiconductors or deposited dielectrics) to form a very high q-factor cavity that brackets a high-gain medium, typically semiconductor quantum wells. Because of the narrow bandwidth associated with the high-q dielectric mirrors, they must be tuned very carefully to the laser emission wavelength of the quantum wells. This fine-tuning of the wavelength makes the device very sensitive to operating temperature. The design of the VCSELs was part of a previous project and their structure and fabrication are not discussed in this report.

Stable operation of high power or temperature sensitive devices is accomplished by the use of an appropriate heat sink. For this project, a very specific heat sink and a careful bonding technique are necesary to integrate the VCSEL die into the MCM package. Recent advances in production of chemical vapor deposited diamond films have increased their use as an affordable heat sink and it has been chosen as the heat sink material for this project because of its very high thermal conductivity, (10-20 W/cm·K vs. ~0.5 W/cm·K GaAs) and electrical resistivity (>10^9MΩ-cm). For this application a heat sink capable of transmitting microwave frequency signals was also required so both the thermal and electrical properties of the diamond are being utilized. The optoelectronic die is a small chip, 900μm x 900μm, and has 36 contact pads on its surface, some on top of active devices. This limits the available packaging techniques to flip-chip bonding, shown in Figure 1. This figure shows a heat sink material that is patterned to match the flipped die and has solder pads where the two chips will come into contact. The die is 'flipped' and aligned to the sink and then the two are brought into contact. The two chips are held in position at a specified pressure and heated to the melting point of the solder. When cooled, the two parts are electrically and metallurgically bonded.



Chip to be bonded Heat sink or 'carrier'



Flip-chip mounted package with light emitting vertically

Figure 1. Flip-Chip Basics

The patterning of the CVDD used in this project was a technological challenge. Because of the relatively high price of CVDD (~$27 per 1.5mm x 1.5mm die), only small sizes were financially feasible. The small size of the die required the development of a set of procedures that allowed patterning of the metallization on small substrates. An encapsulation technique and a tiling technique addressed the problem by making the wafers have an effective size that was compatible with standard processing procedures.

The final aspect of the optical interconnect project addressed in this program was the actual bonding of the VCSEL die onto the patterned CVDD heat sink. This was accomplished by evaporating an appropriate solder material — in the initial tests this has been pure indium — onto the contact pads and then using a flip-chip bonding machine[1] located at RL/OCPB to assemble the pieces.

[1]The M-8A Flip Chip Aligner Bonder is made by Research Devices, 121 Ethel Rd. West, Piscataway, NJ 08854.

Heat sinks were fabricated and a series of practice bonds were performed with appropriate 'dummy' chips and a reliable bonding process was developed. Real devices and sinks were bonded and the resulting package was successfully electrically tested. (An SEM picture of a bonded device is included later in this report.) Another related goal was to expand the metal deposition capabilities of a laboratory at Cornell University to include indium- and tin-based solders. The evaporator is part of an open facility that is currently used by several members of the RL/OCPB staff.

III. Methodology/Results

This section will deal with the three aspects of the packaging of the VCSEL wafers: (1) fabrication of the CVDD heat sinks, which includes the choice of metallizations and patterning techniques, (2) the deposition of solder for the flip-chip application, and (3) the procedures for flip-chip bonding the die onto the sink.

Chemical vapor deposited diamond wafers are available from several manufacturers in a wide range of standard sizes, with the option to have almost any size and shape laser-cut to order, and with several standard or custom metallizations. The VCSEL array die measures 900 μm on a side and needs only about 50 μm clearance from the edge of the mounted die to allow DC and microwave probing. However, to make the diamond more manageable, and to allow for some processing tolerances at the edges of the wafer, a size of 1.5mm x 1.5mm was chosen for the CVDD. It was later determined that the size of the diamond could be enlarged to a more manageable size without any significant increase in material price and 2.5mm x 2.5mm chips have been ordered. These larger dies will also be compatible with standard wire-bonding techniques.

Standard metallizations for CVDD, Ti/Pt/Au and TiW/Au, were not used for this project because of the necessity to wet chemical etch a pattern into the metallization. Pt and TiW are difficult to etch using wet chemistry and only very late in the project were recipes for wet chemical etching both of these materials

found. A simple Ti/Au metallization was chosen because there are simple, reliable, wet etching techniques for both titanium and gold. The parts were specified to be metallized on one side only with 2000Å Ti/2000Å Au. These pieces were mounted and then patterned using photolithography.

IIIA. Mounting

Photoresist is spun onto wafers in order to obtain very uniform thicknesses over large areas, but there is always some kind of edge effect. Due to the viscosity of the photoresist, a bead on the order of .5mm to 1mm thick forms at the edge of the wafer. This is typically of no consequence because processing is usually done in the middle of a large wafer, but for the heat sinks used in this project, the edge bead covers the whole wafer. Two techniques, encapsulation and tiling, were used to make the small pieces of CVDD look like much larger pieces of material. This allows a uniform spreading of the resist without any edge bead effects. The series of pictures in Figure 2 show the encapsulation process.

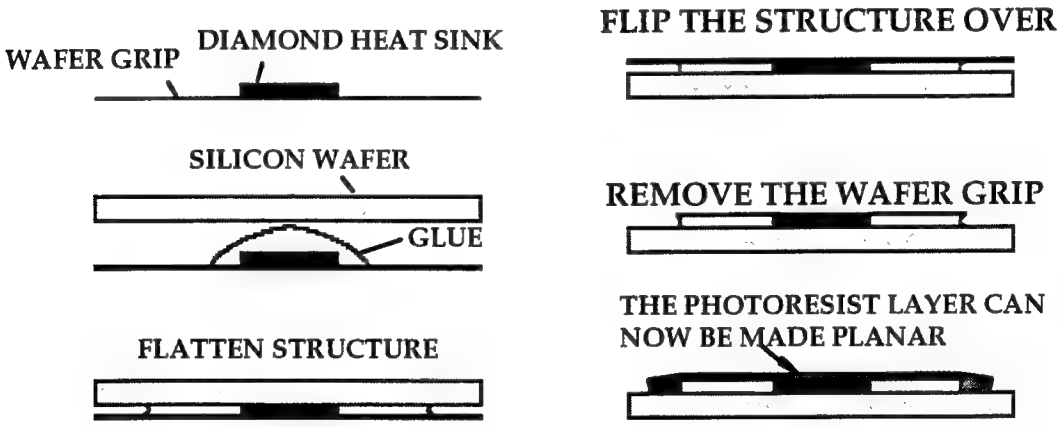


Figure 2. Encapsulation technique

The encapsulation process involved mounting a heat sink in a layer of glue on a carrier with the goal of keeping the top surface of the assembly flat and parallel to the top surface of the heat sink. This gives a large, flat surface on which to spin photoresist. Squares of silicon wafers from 1/2" to 1" on a side were used as carriers, while the best material that served as the mounting 'tape'

was a material called Wafer Grip[2]. Wafer Grip consists of a stable plastic backing with a thin layer of wax on one side. The material is placed on a hot-plate, wax side up, and heated to approximately 90°C, at which point the wax melts. The chip is then placed face down into the soft wax layer and then both are removed from the hot plate and allowed to cool. Many different types of tape, ranging from common household celophane tape to industrial Kapton tapes with special glues, were tried with limited success. The stiff backing plastic of the Wafer Grip kept the final surface of the encapsulation more planar than any of the tape schemes did. A few drops of glue were placed on top of each diamond chip and a silicon square was placed on top of the glue. A small weight was used to keep the assembly flat and to squeeze out all but a thin layer of glue and the assembly was allowed to cure overnight.

After the drying stage, the assembly was re-heated on the hot plate to soften the wax, and the Wafer Grip was pulled off. The encapsulation was then cleaned using a small amount of trichloroethylene and a cotton swab. (Methylene chloride can be substituted for the trichloroethylene and the same results will be obtained.) Because this does not completely clean the wafer (a small film of the wax will remain) a standard oxygen plasma etch was used to prepare the wafer surface for later processing. The chips can be easily removed from the glue encapsulation using acetone. Soaking for several hours or overnight is necessary at times, but all of the samples that were mounted in superglue were recovered after a few hours of soaking. Many of the other glues tried were insoluble in all of the solvents used in standard semiconductor processing facilities.

IIIA2. Tiling

The other mounting process, tiling, met with much more success, but required a large amount of practice and a very steady hand. This process consisted of using a standard photolithography spinner to dispense a gluing compound onto a piece of silicon wafer. Then, the diamond chips were dropped into this glue and arranged so that they were all flush to each other with no gaps.

[2]Wafer Grip is made by Dynatex Corp., 727 Shasta St., Redwood City, CA 94063.

The trick was to get the wafers together without any glue getting onto their top surface. If spun in too thin a layer, the glue dried too quickly. If it was too thick, then it ran onto the top of the chips. (See Figure 3.) All of the interior wafers in the tiling were then edge-bead free, and the edge wafers were 'sacrificed' for the ability to get good resist profiles on the interior samples.

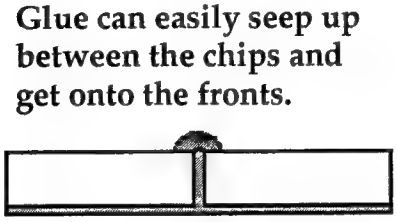


Figure 3. Glue Spreading

The difficult part of this process is getting the proper type and thickness of gluing compound. The solvents in photoresists dry far to quickly to allow any time to mount the samples (and the solvents are typically carcinogenic) and most of the other glues tried were simply too thick to spin onto a wafer. The material that did work was a thick polyimide formulation. (Spin Ciba-Geigy Probimide 287 at 3200RPM for 30 seconds. After positioning the wafers, bake at 90°C for 30-45 minutes.) This material dried slowly enough to allow the assembly to be adjusted and to allow large (>20 pieces) tilings to be formed. At the end of the processing, methylene chloride or n-methyl-2-pyrrolodinone was used to soak the CVDD off the silicon carrier. All of the sinks fabricated in this project were mounted using this tiling technique.

Both techniques, encapsulation and tiling, are very promising and with the proper combination of adhesives and solvents, both processes can be made to work very well. In an industrial application, where the initial material cost is only a small part of the total price of the project, much larger pieces of diamond could be used. Then, standard lithography techniques could be used from the start and after processing, the dies could be cut apart using a diamond saw or a high power laser dicing system.

IIIA3. Processing

Although the heat sink pattern was designed previously, its transfer into a metallization pattern on the heat sinks was implemented under this contract. The pattern consists of a set of coplanar waveguides (CPW) that extend contacts from the flipped surface of the VCSEL to the outside of the die. Coplanar waveguides are a configuration of microwave transmission lines that allows both DC and microwave signals to propagate to the device. The final waveguide structure needs to have both a thick gold metallization and a solder of an appropriate thickness deposited onto the contact pad areas. The thick metallization is necessary to minimize microwave losses in the transmission lines and to interact with the solder during the bonding process. Because of the specified custom metallization on the heat sinks, a thick gold layer would lead to severe processing problems for the manufacturer, so the thick metal had to be realized at our facility.

The process steps to make a thick metallization layer underwent several evolutions over the summer with two techniques completed and a third partially completed. The first technique was to simply use standard lift-off technology with thick photoresist to obtain a 1.5 to 2μm thick gold metallized pattern on top of the thin Ti/Au initial metal. The base gold is then patterned with TFA gold etch with the deposited gold acting as a sacrificial mask. The plan was to remove 2000Å of the lower metallization and 2000Å off the top and sides of the deposited metal. The ratio of thicknesses of the gold underlayer and overlayer should have allowed a large margin for overetching while retaining an adequate thickness of evaporated metal, but the evaporated gold film was very porous and showed selective etching. There was also a problem with non-uniform etching of the underlayer, possibly due to photoresist residue. The exposed titanium etched very easily and reliably in a mixture of HF:DI::1:50, which etched the 2000Å of Ti in about 11 seconds. This etch was used for all subsequent titanium etches. The results of this first process were not very promising, but it warrants further study because of its ease of processing (one lithography step) and potentially very well controlled tolerances in pattern size.

The second technique used to obtain a thick gold layer was electroplating. The initial metallization was etched into the correct pattern with an extra piece of metallization left intact in order to electrically connect all of the pattern. The parts were then to be plated to the desired thickness and patterned with photolithography to remove the extra metallization ring. In order to try this technique, dummy sinks were fabricated on semi-insulating GaAs wafers which were glued onto glass slides. A special jig was fabricated out of Teflon to hold the glass slide and it included a nylon screw to position a BeCu electrode onto the corner of the die. The process of mounting the die was not very difficult with a little bit of practice, but because of agitation during plating, the sample almost always had a tendency to slide out from underneath the electrode, making most of the plating runs useless. This technique seemed very simple on the surface, but it does require two photolithography steps and a special mounting in order to attach an electrode for the plating process. Effort on this metallization technique was stopped when it was decided that the process was neither reliable nor repeatable.

The third technique for depositing a thick metal was to continue the process that the manufacturer used for depositing the initial metal, sputter deposition. Sputter deposition is a very efficient way to deposit material because, unlike evaporation, almost all of the sputtered target material is deposited on the sample. Since the material in question is gold, this is an important consideration. There are facilities at Cornell University to deposit gold by sputter deposition and using this process 2μm of gold was deposited. (There was not time to do any further processing on the wafer, but work will continue on this process under separate funding.) The next step in this process is to mount (by encapsulation or tiling) and pattern (by photolithography and etching) the thick gold layer, modifying the pattern dimensions to include an appropriate undercut of the gold lines during etching. This process is inviting because many wafers can be metallized at once, using a very time-, money-, and gold-efficient technique, and only one lithography step is used. One problem is that there is significant heating of the wafer during the sputter deposition, polymerizing the photoresist that is used to 'glue' the wafers down and making it very resistant to solvents. Soaking

in n-methyl-2-pyrrolodinone and acetone overnight removes most of the resist and a strong oxygen plasma is expected to adequately clean the backs of the wafers.

IIIB. Solder

All three of the techniques for making thick gold lines provide identical starting points for the deposition of solder. The deposition of solder is done using standard photolithographic lift-off processes and thermal or e-beam evaporation of the solder. The choice of solder for this project is not a rigid as that for an industrial project, but several issues were considered. The solder had to have a melting point compatible with the flip-chip bonder at RL/OCPB and it has to be compatible with the metallizations on the heat sink and on the VCSEL die, both of which are gold. The bonder was adequate to melt all solders considered for this project, but bonding to gold eliminated many standard solders because they tend to 'scavenge' gold from the contact pads and the chips can simply fall off. Indium-lead based solders with minimal scavenging properties were considered, but rejected because of the toxicity issues of handling and evaporating relatively large amounts of lead. Pure indium is used in some applications as a solder because of its low melting point (156.7°C) and good wetting properties, but has the unfortunate property of scavenging gold. A eutectic mixture of gold and tin (80/20) with a melting point of 284°C was also considered, because it is a very important industrial standard for moderate temperature soldering applications in which gold contacts are used, but turns out to be very expensive. The availability of indium and its ease of handling, evaporation, and melting, led us to use it as a solder.

A third issue of solder choice, reliability, is one faced by industrial applications. Thermal cycling of the package causes large stresses to be placed on the chip, the solder bonds, and the heat sink, and a certain amount of flexibility of one or more of these parts is necessary. The GaAs wafer is very brittle and the CVDD is very strong, so the pliable material must be the soldering alloy. Indium is a very soft metal, but the small amounts used in this application will not allow much room for flexibility of the final structure. The gold-tin solder tends to be brittle, but has the important quality of making very good bonds to gold. Because

these samples are not going to be subjected to extremes in temperature, and the die is very small, many of the reliability criteria have been relaxed for this project. The short term mechanical and electrical stability of the devices were the immediately important parameters so indium will continue to be used and gold/tin has been ordered for future work requiring higher temperature post-bonding processes.

The solder metallization of a carrier that is used for flip-chip bonding is typically relatively thick in order to allow for differences in thermal expansion of the semiconductor and the carrier, but a thick solder is not possible in this application because the bonding occurs directly on top of devices, not on bonding pads next to the devices. Figure 4 shows an assembly drawing of a

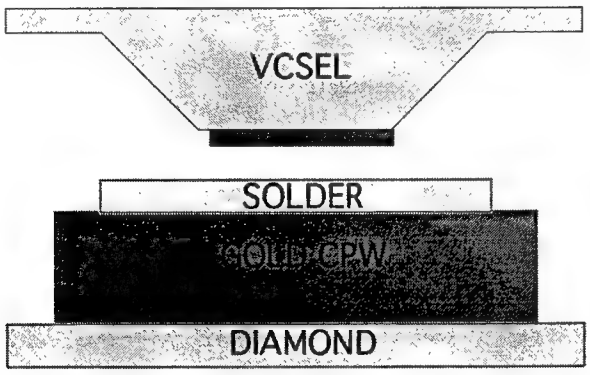


Figure 4. Cross Section of Bond Pad

single bond pad. The VCSELs were fabricated using wet chemistry to define the mesas and have a metallization on top of the mesas that is capped with a thick gold layer. Any excess solder that might squeeze out from between the pads could easily spread onto the mesas and short the devices, so solder thicknesses less than 2μm were used. Because this is not a traditional soldering application, and such small amounts of solder are being used, scavenging of the gold might not be an important issue. (All solder evaporations were done at Cornell University.)

Several evaporations of indium were made to determine deposition conditions for good quality films. It was found that the evaporation had to be done at a moderate to slow rate (10-40Å/sec) in order to keep the sample

temperature low. High evaporation rates and high sample temperatures cause macroscopic roughness and a very noticeable discoloration to white, yellow, or brown. A good film a few thousand angstroms thick should be very shiny and a film a couple microns thick should still be relatively shiny. The test indium films were then patterned and melted in the flip-chip bonder to verify that the films were pure enough to reflow at or near the expected melting point of 157°C, and melting was observed by 160°C for several different evaporation conditions. A completed heat sink is shown in Figure 5.

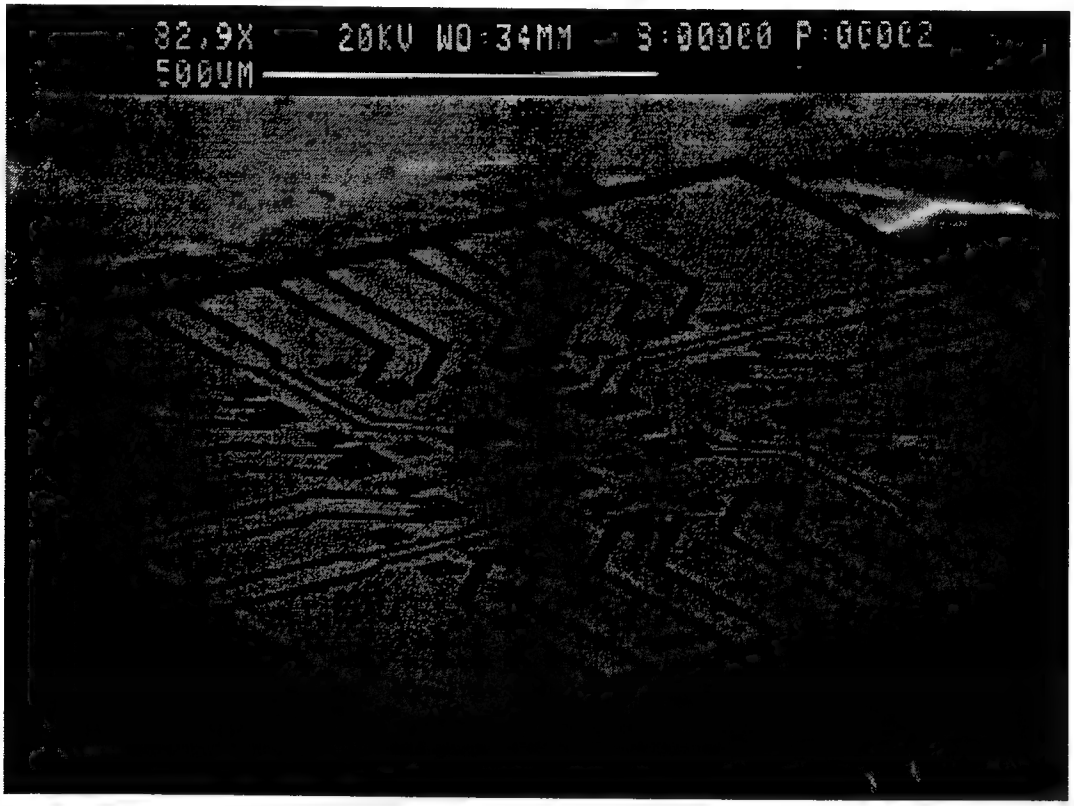


Figure 5. Completed Diamond Heat Sink

In order to continue this and other bonding work, the ability to deposit a potentially large range of solders is required. Because of the toxic and/or contaminating nature of the material in the solders (lead, indium, tin, etc.), there are not very many facilities that allow solder evaporation. During this project, a used but still-working thermal evaporator became available for dedicated use in this project. A couple of weeks of time at the end of the summer were spent modifying the basket mounting assemblies and substrate holder for this

evaporator. Because of delays in getting necessary parts fabricated in machine shops or delivered from outside vendors, the evaporator was not completed, but work will continue under other funding. (Because of the length limits of this report, drawings of the evaporator parts are not included, but may be obtained from the author.)

IIIC. Bonding

Test structures of heat sinks with base metallizations of indium on glass were fabricated and reject VCSEL dies were mounted on them in order to test the flip-chip bonder alignment accuracy and to develop recipes for the bonding process. Even though glass is a poor thermal conductor, it was used so that alignment could easily be checked through the substrate. A problem occurred during the evaporation of the indium onto the glass, so solder reflow and bonding didn't occur, but the solder was thick enough, and the indium was sticky enough, that the die did attach to the wafers. This was actually a fortunate result because the die came off the sinks very easily after most bonding attempts. This allowed the planarity and alignment of the die and sink to be measured by looking at the uniformity and position of the depressions made in the solder by the bond pads of the VCSEL die. These first tests of the flip-chip bonder showed that the planarity of the two chips was almost perfect, and due to the accuracy of the bonder, the alignment was checked to be well within the tolerances required by this project. (±5μm would be considered a perfect bond and ±10μm could be tolerated. The machine is specified to have better than ±2μm post-bond positioning accuracy) Some of these initial devices did physically bond enough to give realistic mechanical stability and electrical conductivity and LED emission was demonstrated from many of the devices. Subsequent evaporations onto glass, GaAs, and CVDD resulted in well-behaved indium films.

A several step bonding process sequence was developed over the course of several days using an assortment of materials for the top and bottom wafers, including diamond heat sinks, glass, silicon, and GaAs VCSEL wafers. It was found that the material compositions didn't affect the times or temperatures of

the process. The mass of the heated chucks was so much larger than the samples that their varying thermal conductivities and thicknesses had no effect on the process. Tests with small fragments of pure indium showed that there was significant deformation of the fragment even well below the melting point, and that as soon as the melting point was reached, the fragment could not support any applied pressure and was immediately flattened between the chucks. This resulted in a large amount of indium squeezing out from between the chucks (protected with small pieces of silicon wafers). This reinforces the guideline for using a thin layer of indium as solder.

The bonding sequence was suggested by the instructions provided with the flip-chip aligner bonder, but the times and temperatures were determined experimentally for this choice of solders and dies. Figure 6 shows a diagram of

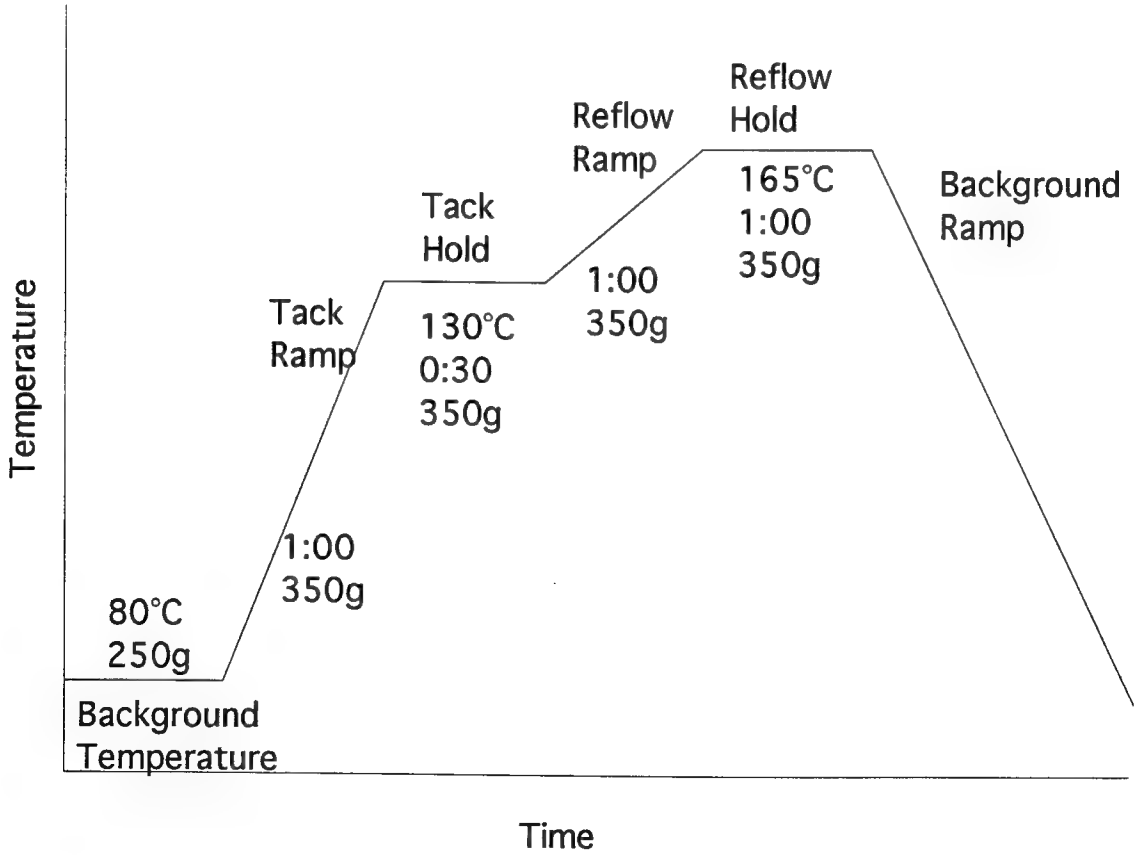


Figure 6. Bonding Parameters

times and temperatures for the flip-chip bonding process. The temperatures were chosen based on the use of indium as solder while the hold times were almost arbitrary. Times were kept short based upon the results of melting the indium chunks and the ramp times were determined by the heating cycles of the bonding apparatus. Figure 7 shows an assembled VCSEL die and heat sink. This and

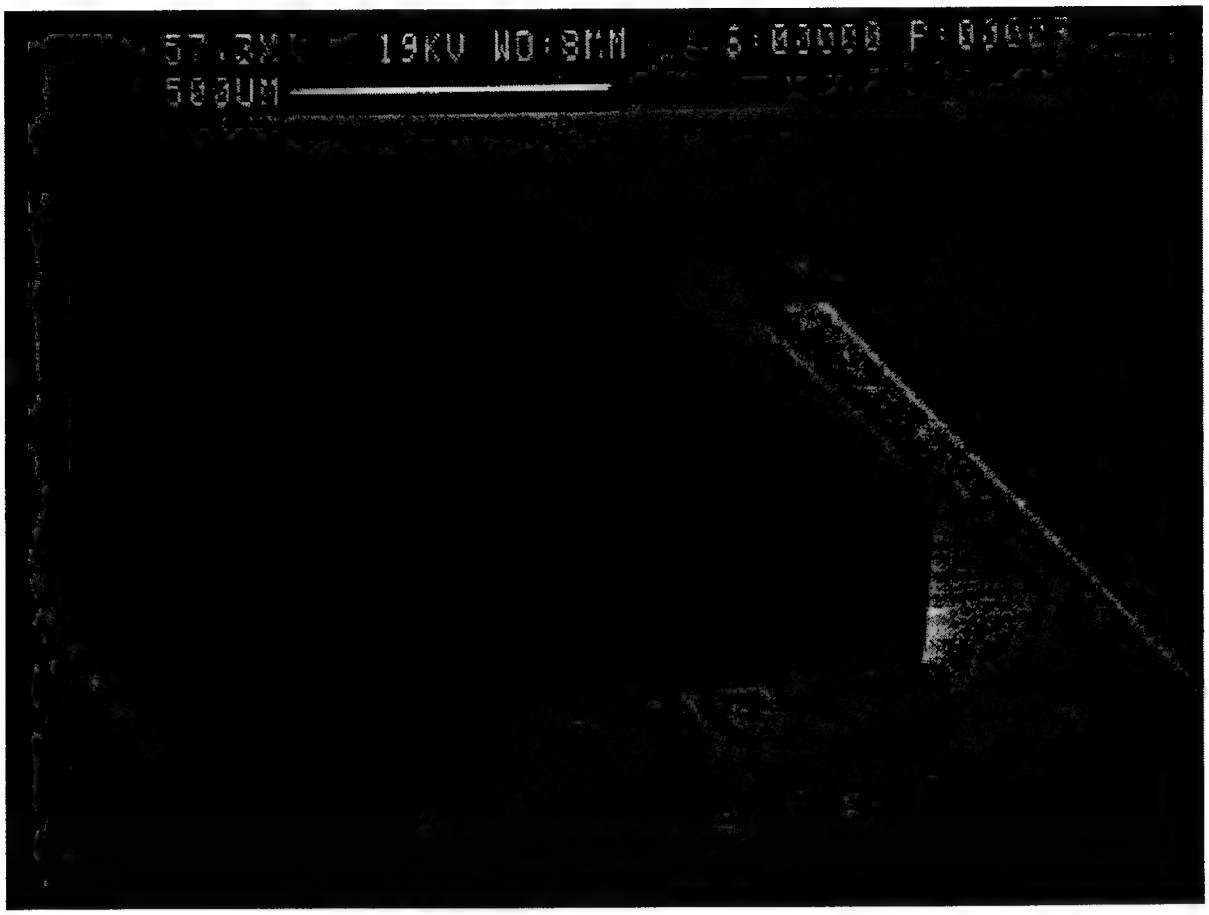


Figure 7. Completed VCSEL die on heat sink

several other assembled packages were subjected to a drop test. To check the mechanical stability of the bonds, the assembled chips were dropped from one to two feet onto lab tables covered with a layer of clean wipe. All devices bonded with the new recipe survived all drops. These devices were then electrically tested for diode characteristics and the VCSELs tested for light emission. The bonding process was successful and all bonded devices emitted light.

Summary/Conclusions

The use of optoelectronic devices in high performance systems requires special attention to thermal control and stability. In this study two techniques (encapsulation and tiling) for fabricating heat sink patterns on small CVDD chips were developed. These CVDD chips were then used as heat sinks for vertical cavity surface emitting laser arrays. A recipe for flip-chip bonding the VCSELs onto the sinks was developed and electrically and metallurgically bonded devices were presented. Work was also done on retrofitting an evaporator to allow the use of different types of solder in future applications.

MULTISENSOR-MULTITARGET DATA FUSION USING AN S-DIMENSIONAL SLIDING WINDOW ASSIGNMENT ALGORITHM

Robert L. Popp
Ph.D Candidate
Department of Computer Science and Engineering
rpopp@sparc0.brc.uconn.edu

University of Connecticut
UBox 155
Storrs, CT. 06269

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

and

Rome Laboratory

September 1994

MULTISENSOR-MULTITARGET DATA FUSION USING AN S-DIMENSIONAL SLIDING WINDOW ASSIGNMENT ALGORITHM

Robert L. Popp
Ph.D Candidate
Department of Computer Science and Engineering
University of Connecticut

Abstract

The tracking process as generally practiced today consists of the interrelated functions of association and estimation. Association is the decision process of linking observations (measurements) of a common origin (target). A set of linked observations can then be statistically filtered to estimate the states of targets. The central problem in the multisensor-multitarget data fusion process is that of data association, that is, the problem of determining from which target, if any, a particular measurement originated. Measurements originating from a particular target can then be fused to estimate the states of targets not directly measured by sensors. Furthermore, in dense multisensor-multitarget scenarios, data association is the most computationally expensive aspect of tracking. Mathematically, the data association problem for tracking can be formulated as a multidimensional assignment problem, for which is NP-hard, and therefore intractable when the degree of dimensionality is greater than 2. However, the S-dimensional assignment algorithm (*S* refers to the number of dimensions) developed by Pattipatti [9,10] and Deb [4-6], is a suboptimal Lagrangian relaxation-based technique that is an efficient and polynomial-time solution to the generalized multidimensional assignment problem.

To exacerbate the problem, when the degree of data association is across too few scans, misassociations are more probable and thus become more problematic. As a result, these misassociations inevitably will lead to poor track performance, loss of track, inaccurate estimated target position, and tracking errors far worse in reality than those predicted by the estimation scheme. However, because of the lack of any a priori estimates of target states, reliable track initiation may improve if the data associations are optimized over multiple scans. Moreover, in track extension, misassociations can be reduced significantly if the data association is optimized over multiple scans [4]; any misassociations in earlier scans will result in a subsequent degradation in solution quality; hence, optimizing the solution of S scans will significantly reduce the chance of misassociation. The near-optimal S-dimensional assignment algorithm described in [4-6] is highly efficient, recursive, and ideally suited for a multiscan sliding-window implementation.

The purpose of this paper is to demonstrate the use of the "sliding" multiscan windows approach in solving the multidimensional assignment problem for both track initiation and track extension. As part of AFOSR funded research and the current research effort for Rome Laboratory, the University of Connecticut has developed several data association and estimation algorithms for actual tracking problems of interest to the Air Force. One such algorithm developed by Pattipatti [9,10] and Deb [4-6] is the near-optimal, highly recursive, polynomial time S-dimensional assignment algorithm. In this work, as part of AFOSR's Summer Research Program, we implemented and tested (via an actual non-stressing multisensor-multitarget data set) a multiscan sliding-windows tracker utilizing the S-dimensional assignment algorithm as our data association component. The overriding objective was to demonstrate the use of a (sliding) multiscan window approach in addressing the problem of efficient and reliable data association, over time, for track initiation and extension.

MULTISENSOR-MULTITARGET DATA FUSION USING AN S-DIMENSIONAL SLIDING WINDOW ASSIGNMENT ALGORITHM

Robert L. Popp

Introduction

A surveillance and tracking system consists of a network of geographically and functionally distributed sensors (e.g., L-band, S-band, C-band, infrared). Besides estimating the target state variables (e.g., position, velocity, acceleration, etc.), one of the central objectives of such a system is to detect an unknown number of targets in its field of view and to associate the target detections with either pre-existing formed tracks, or to initiate a new track for the target; moreover, this must all be accomplished in an environment where sensor measurements are contaminated by noise, spurious observations created by noise and clutter are present, and occasionally, the sensor misses a target detection.

Another important fundamental objective of surveillance is to produce highly accurate target and track identification in real-time, even in dense target scenarios and in regions of clutter and high track contention. The utilization of multiple sensors has the potential to greatly enhance track identification. Moreover, for multiple target tracking, the processing of multiple scans simultaneously yields high track identification. However, to achieve such track identification in real-time, one must solve an NP-hard data association problem. To date, the only known method for solving this problem optimally in both multisensor data fusion and multitarget tracking is via branch and bound algorithms (e.g., Multiple Hypothesis Tracking (MHT)). However, with the emergence of fast, near optimal algorithms in solving the multidimensional assignment problem (i.e., Lagrangian relaxation based algorithms), the ability to solve this track identification problem, suboptimally, but with relatively high accuracy, in real-time now exists [11-15].

As indicated in [1-3,17], in the development of the tracking algorithm, as currently practiced, four interrelated processes are significant: **i)** state variable model selection used in the target motion and sensor measurement representations, **ii)** score function evaluation for the candidate measurement-target associations, **iii)** measurement-target association consistency determination, and **iv)** target states estimation. As indicated in **ii)** and **iii)**, central to the tracking process is data association. Association is defined in our context as the decision process of linking measurements of a common origin (i.e., target or false alarm) such that each measurement is associated with only one origin. The predominant problem in multisensor-multitarget tracking is the data association problem, that is, partitioning the scan observations into tracks and false alarms so that an accurate estimate of the true tracks can be recovered. Moreover, when the multisensor-multitarget scenario is dense, the data association problem becomes the most computationally

expensive aspect of tracking. However, when the degree of data association is across too few scans, misassociations are more probable and thus become more problematic. As a result, these misassociations inevitably will lead to poor track performance, loss of track, inaccurate estimated target position, and tracking errors far worse in reality than those predicted by the estimation scheme. But if we optimize the data association over multiple scans, and thus potentially reducing the possibility of misassociations, reliable track initiation and extension can be realized [4]. This is because any misassociations in earlier scans will result in a subsequent degradation in solution quality; hence, optimizing the solution for S scans has the potential to significantly reduce the chance of misassociation.

The use of combinatorial optimization in multisensor-multitarget tracking is not new and dates back to the pioneering work of Sittler [16] and Morefield [8], and in more recent years, Bar-Shalom, et al. [1-3], Deb, et al. [4-6], Pattipatti, et al. [9-10], and Poore, et al. [11-15]. Although combinatorial optimization is the natural framework for the formulation of these data association problems, the corresponding techniques have long been considered computationally too expensive for real-time applications. The corresponding optimization problems for multiscan processing, which are in this work formulated as multidimensional assignment problems, are NP-hard. To further appreciate the difficulties, one only need consider the tradeoffs associated between the two predominant methodologies in multisensor-multitarget tracking: single-scan processing and multiscan processing. In the former case, one essentially extends tracks a single scan at a time via, for example, a two-dimensional assignment algorithm to associate observations to tracks. This was the approach in [17] for the exact same data set as used in this work. Many of the previous works in multisensor-multitarget tracking are based on a single-scan processing methodology [17], which does, in general, provide for a real-time characteristic, but may potentially lead to a large number of partial and incorrect assignments, particularly in data stressing cases (e.g., cluttered, noisy environments, crossing, splitting, and merging tracks), and thus incorrect track identification. The primary difficulty with this approach is the lack of information in single-scan processing to properly partition the observations into tracks and false alarms [14]. Moreover, because there is no history in the data association portion of the tracking algorithm, the accuracy of the estimation portion (e.g., Kalman filtering) becomes all the more important. However, it should be noted that due to the non-stressing nature of the data set provided by Rome Labs, the work done in [17] did not suffer from these difficulties. In the latter case, the processing of multiple scans at a time to a large extent resolves these difficulties, primarily because of the built in history in the data association part of the tracking algorithm. However, as was previously mentioned, the multiscan processing approach is computationally expensive for real-time use. Implementation of the tracking algorithm on a high-performance (multiprocessor) computer thus becomes important when considering the multiscan processing approach.

As part of the current research effort for Rome Laboratory, the University of Connecticut has developed several data association and estimation algorithms, developed as part of AFOSR funded research, for actual tracking problems of interest to the Air Force. In past research [17], one of the overall

objectives was the development of the maximum likelihood formulation of the data association problem and solving the resulting problem using a sliding window 2-dimensional assignment algorithm (i.e., auction algorithm). In that work, established target tracks are associated with the new measurements from the latest scan, thus, a 2-dimensional assignment problem results. Given the non-stressing target measurement data set provided by Rome Labs, the near optimal association obtained using the 2-dimensional assignment algorithm made it unnecessary to use the more general and robust S-dimensional assignment algorithm developed in [4-6]. As was noted, in order to fully utilize the advanced features of the the S-dimensional assignment algorithm, it should be used in more complex target scenarios, for example, in situations where target tracks split, merge, and cross. Hence, the objective of this work follows:

- development of the maximum likelihood formulation of the data association problem for the actual tracking scenario.

- implementation of a multiscan sliding window tracker utilizing the S-dimensional assignment algorithm developed in [4-6] as the data associator; furthermore, this implementation should be in such a way so as to allow for an easy installation into the Rome Laboratory Fusion Tracker II testbed.

- generation of simulated target scenario stressing data sets consisting of a cluttered, noisy environment with crossing, splitting, and merging tracks.

- performance analysis, in terms of track identification and extension quality, wall-clock time necessary in the multidimensional association, and its relative performance compared with the one scan at a time approach as demonstrated in [17].

- development, implementation, and analysis of the S-dimensional assignment algorithm on a high-performance (multiprocessor) computer, in particular, such performance analysis when the multisensor-multitarget scenario is dense.

Without giving any discussion here, in Figure 1 we provide an overall block diagram, as it pertains to the tracking process, of our implementation of the multiscan sliding-window tracker via the S-dimensional assignment algorithm. Later in Section 6 we provide details and some initial results on our implementation.

The purpose of this paper is to demonstrate the use of the "sliding" multiscan windows approach in solving the multidimensional assignment problem for both track initiation and track extension. As part of AFOSR funded research and the current research effort for Rome Laboratory, the University of Connecticut has developed several data association and estimation algorithms for actual tracking problems of interest to the Air Force. One such algorithm developed by Pattipatti [9,10] and Deb [4-6] is the near-optimal, highly recursive, polynomial time S-dimensional assignment algorithm. In this work, as part of AFOSR's Summer

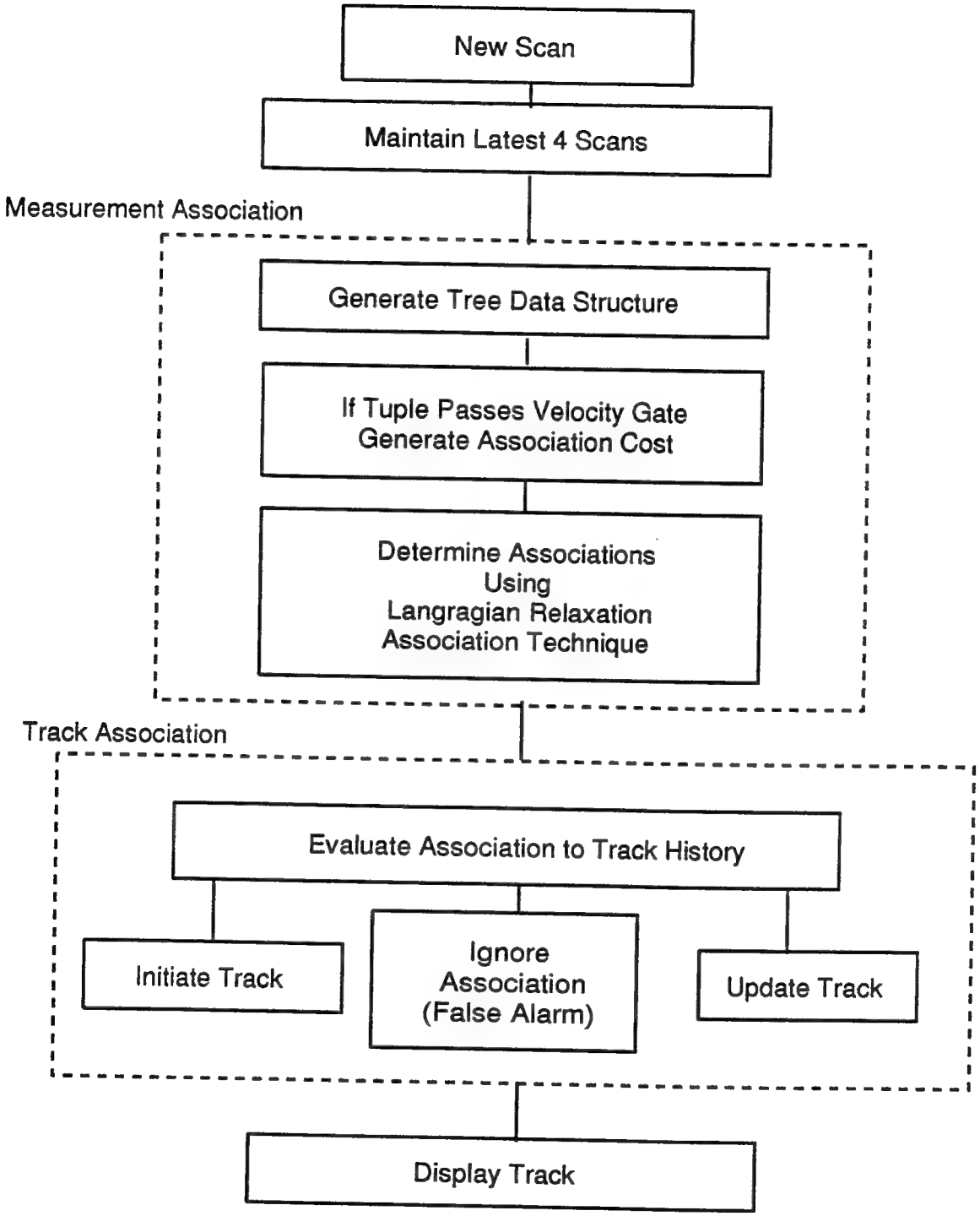


Figure 1. Block diagram of multiscan sliding-window tracker via the S-dimensional assignment algorithm.

Research Program, we detail our research efforts as pertaining to the first two items of the above list, and, for further research efforts, we delay until later any results associated with the latter three list items. In this work, we implemented and tested (via an actual non-stressing multisensor-multitarget data set) a multiscan sliding-windows tracker utilizing the S-dimensional assignment algorithm as our data association component. The overriding objective was to demonstrate the use of a (sliding) multiscan windows approach in addressing the problem of efficient and reliable data association, over time, for track initiation and extension.

The remainder of this paper outlines as follows. In Section 2 we provide a description of the data set provided by Rome Labs for this particular research. This is followed by some statistics for the data set in Section 3. Section 4 provides a brief introduction to the S-dimensional assignment algorithm, and Section 5 a mathematical formulation of the S-dimensional assignment problem underlying this research effort. In Section 6 we provide some of the implementation details and initial performance results, and follow this with future research directions and conclusions.

2. Multisensor-Multitarget Data Set

In this section we provide a description of the raw scan measurement data to be used in the following work. The measurement data consists of scans from two L-band FAA radars located at two locations in New York, one at Remsen, and the other at Dansville. The raw measurement data file consist of 55 scans arriving from the Remsen and Dansville radar sites asynchronously, at approximately every 10 seconds apart. Each of these scans contains a number of radar beacon or radar skin returns. Each of the skin returns consist of a time stamp, a slant range and azimuth angle measurement. For cooperative targets, a beacon return is obtained, which provides, beside the time stamp, slant range, and azimuth, a target id number (squawk) and altitude measurement. In our work in the S-dimensional assignment algorithm implementation, for the observability of a target, a full measurement of its position is necessary. Only beacon returns provide such measurements. Note that the availability of the measurement id provides a means for evaluating the performance of the overall S-dimensional assignment algorithm with respect to both track initiation and extension. The number of detections in each scan used in the S-dimensional sliding window assignment algorithm ranges from 22 to 43.

Because the scenario of interest cannot be modeled as linear both due to the nonlinearity of the measurement equations as well as the large geographical area involved, it was necessary to use a spherical earth model in place of a flat earth approximation. Furthermore, because all scan measurements were assumed to be either a target position or false alarm in the sensor(s) field of view space, a 3-dimensional Cartesian coordinate frame of reference was needed. Thus, with each sensor detection, there is an associated measurement, which we transform from its raw representation to an Earth Centered Rotating (ECR) Coordinate system, which, by definition, is defined to be fixed with respect to the earth with its

origin at the center of the earth. The xy-plane of this coordinate system is the equatorial plane of the earth with the x-axis aligned to the 0° longitude and 0° latitude and the z-axis aligned with the North Pole (the rotational axis of the earth). Note that the longitude is measured positively from the x-axis toward the y-axis. We represent a position in ECR coordinate system by, say $(\lambda_l, \delta_l, h_l)$.

Given a sliding window of say S = 4 scans, once the data association is made, we convert the measurement from ECR to an ECR Geodetic Coordinate (ECRGC) system for purposes output display in Rome Labs Fusion Tracker II testbed. The ECRGC coordinates are the longitude, latitude, and altitude of a point in ECR coordinates. The relationship between the orthogonal coordinates, say Z_o^R, in ECR, and the geodetic coordinates, say $(\lambda_L, \delta_L, h_L)$, is based on the Oblate Earth model which assumes the earth shape is an ellipsoid which is rotationally symmetric about the z-axis.

$$Z_o^R = \begin{bmatrix} \left\{ \frac{R_E^2}{\sqrt{[R_E cos(\delta_R)]^2 + [R_p sin(\delta_R)]^2}} + h_R \right\} cos(\delta_R)\ cos(\lambda_R) \\ \left\{ \frac{R_E^2}{\sqrt{[R_E cos(\delta_R)]^2 + [R_p sin(\delta_R)]^2}} + h_R \right\} cos(\delta_R)\ sin(\lambda_R) \\ \left\{ \frac{R_p^2}{\sqrt{[R_E cos(\delta_R)]^2 + [R_p sin(\delta_R)]^2}} + h_R \right\} sin(\delta_R) \end{bmatrix}$$

where R_E is the equatorial radius and R_p is the polar radius.

3. Sensor Measurement Statistics

The measurement noises of the sensors are assumed to be zero-mean white with the following variances: for the range measurement, the resolution cells are

$$\Delta r = \frac{c\tau}{2},$$

where $c = 3.\ 10^8 m/s$, and τ is the pulse width (6 us for Remsen and 1.8 us for Dansville). We assume the range measurement is uniformly distributed in each resolution cell, hence, the standard deviation for the range measurement noise is

$$\sigma_r = \frac{\Delta r}{\sqrt{3}},$$

which yields 0.5196 km = 0.2806 nmi for Remsen and 0.1559 km = 0.0842 nmi for Dansville. The azimuth measurement noise standard deviations were taken as

$$\sigma_\theta = 2.618\ mrad = 0.15°$$

based on FAA data. No other specific information was given. Finally, the altitude measurement noise, based on FAA data, is

$$\sigma_h = \sqrt{50^2 + \frac{100^2}{12}} = 57.7350 ft = 17.5976 m.$$

The following values have been used for the sensors' probability of detection P_{D_s} and the surveillance region volume for the region in which a candidate false alarm is uniformly distributed:

$$P_{D_s} = 0.999$$
$$\Psi_s = 1000 km^3.$$

The estimated position and velocity for target t is a six-element vector $[\mathbf{p_t}, \mathbf{v_t}]$, with components

$$\begin{bmatrix} \lambda_l & - \text{ ECR longitude} \\ \delta_l & - \text{ ECR latitude} \\ h_l & - \text{ ECR north pole} \\ \dot{\lambda}_l & - \text{ longitudal velocity in km/s} \\ \dot{\delta}_l & - \text{ latitudal velocity in km/s} \\ \dot{h}_l & - \text{ north pole directional velocity} \end{bmatrix}$$

4. S-dimensional Assignment Algorithm

In this section, we present a brief introduction to the S-dimensional assignment algorithm and refer those who are interested in more details to the literature [4-6,9,10]. The multidimension assignment problem, a problem inherent in the multisensor-multitarget data fusion process, is NP-hard. Thus, optimal solution techniques requiring unacceptably long periods of time are of little value. Instead, fast solutions, with relatively good accuracy, are more desirable. The S-dimensional assignment algorithm developed at the University of Connecticut [4-6,9,10] is one such solution technique. The S-dimensional assignment algorithm is a polynomial-time, suboptimal solution, relaxation algorithm that iteratively improves the solution while providing a measure of how close the current solution is to the optimal one. The solution approach to the S-dimension assignment problem (S > 2) is to optimally solve a series of relaxed 2-dimension assignment problems via the generalized Auction algorithm, which has complexity $O(n^3)$, where n is the scan list size. If there are k such relaxations, then the S-dimensional assignment algorithm has complexity $O(kn^3)$.

Instead of trying to obtain the optimal solution from just the feasible (primal) solutions, the solution approach in the S-dimensional assignment algorithm is to successively relax the constraint sets r: r = S,S-1,...,3 and append them to the cost function using Lagrangian multipliers $\mathbf{u}_r$. Thus, at stage r =3, the problem has been relaxed to a 2-dimensional assignment problem, which is then optimally solved. Then the feasible (and probably suboptimal) solution to the relaxed 3-dimensional assignment problem is computed, and the Lagrangian multiplier $\mathbf{u}_3$ is updated. Similarly, the r-dimensional assignment problem, r = 4,5,...,S, is successively solved, and the corresponding Lagrangian multiplier vector $\mathbf{u}_r$ is updated. This cycle of relaxing constraints, solving relaxed reduced-dimensional assignment problems and updating Lagrangian multipliers is repeated until all the constraints in the constraint set are satisfied. Since the relaxed assignment problems do not enforce all the constraints, their solutions will produce a lower bound on the actual optimal solution. The intermediate feasible (primal) solutions are all suboptimal, and have costs greater than or equal to the optimal feasible solution. The *duality gap*, the difference between the two, provides a conservative estimate of the accuracy of the suboptimal feasible solution..

5. Mathematical Formulation

For this research, we solve the following problem: there are N non-maneuvering targets in the target scenario. The constant-velocity linear motion trajectory of target t is uniquely determined by its position $\mathbf{p}_t$ and velocity $\mathbf{v}_t$ at some time instant T_t which is not a given, but, instead, extracted from the data. Originating from sensor s, there are $S = 4$ scan sets of n_s, s = 1,2,3,4, measurements, at time instants T_{si_s}, i_s = 1,2,...,n_s. The measurement $\mathbf{z}_{\mathbf{si}_s}$, an ECR coordinate position, either originated from a true target, in which case it is the true target position, $H(\mathbf{p}_t, \mathbf{v}_t, T_{si_s})$ plus some Gaussian noise, $N(0, \Sigma_s)$, or from some spurious source, in which case it is uniformly distributed within the field of view, Ψ_s, of sensor s. In addition, each sensor has non-unity probability of detection, P_{D_s}. Our task is to identify the unknown number of targets and estimate both their initial position (i.e., track initiation) and current position (i.e., track extension), if applicable.

Our solution approach involves the formulation of a generalized likelihood ratio to assign costs to each feasible S-tuple of measurements, and then employ the S-dimensional assignment algorithm of [14] to globally minimize the cost.

To simplify the notation for incomplete measurement-target associations caused by missed detections, we add a dummy measurement, $\mathbf{z}_{\mathbf{s0}}$, to each sensor's measurement set. The likelihood that a S-tuple of measurements, $Z_{i_1 i_2 \cdots i_S}$, originated from target t is

$$\Lambda(Z_{i_1 i_2 \cdots i_S} | t) = \prod_{s=1}^{S} [P_{D_s} p(z_{si_s} | \mathbf{p}_t, \mathbf{v}_t)]^{u(i_s)} [1 - P_{D_s}]^{1 - u(i_s)}.$$

Note the indicator function,

$$u(i_s) = \begin{cases} 0 & \text{if } i_s = 0 \quad \text{denoting a missed detection by sensor } s \\ 1 & \text{otherwise} \end{cases}.$$

The likelihood that they are all spurious, i.e., $t = \emptyset$, is

$$\Lambda(Z_{i_1 i_2 \cdots i_S} | t = \emptyset) = \prod_{s=1}^{S} \left(\frac{1}{\Psi}\right)^{u(i_s)}.$$

Hence, the cost of associating the S-tuple to target t is given by the negative log-likelihood ratio

$$c_{i_1 i_2 \cdots i_S} = -\ln \frac{\Lambda(Z_{i_1 i_2 \cdots i_S} | t)}{\Lambda(Z_{i_1 i_2 \cdots i_S} | t = \emptyset)}.$$

Note that because the likelihood function is a probability density function, one cannot compare the likelihood function value of 2 measurements with that of 3. However, by using likelihood ratios, this comparison is possible because the result is now a dimensionless quantity [1,14].

Since the target position and velocity $[\mathbf{p_t}, \mathbf{v_t}]$ are unknown in the previous equations, we replace them by their least-squares estimates

$$H(\hat{\mathbf{p}}_\mathbf{t}, \hat{\mathbf{v}}_\mathbf{t}, T_{si_s}) = \mathbf{Z_0} + (T_{si_s} - t_0)\mathbf{V_0}.$$

For example, let $\mathbf{z_{si_s}}$, $\mathbf{i_s}$ = 1,2,3,4, be four 3-D ECR measurements, say $[z_{s1\lambda_l}, z_{s1\delta_l}, z_{s1h_l}]$,..., $[z_{s4\lambda_l}, z_{s4\delta_l}, z_{s4h_l}]$, from sensor s (i.e., Dansville or Remsen). The instantaneous target motion, $[\mathbf{Z_0}, \mathbf{V_0}]$, in the ECR reference frame, approximated by the constant velocity linear motion model, is computed by solving the following system of equations:

$$\begin{bmatrix} z_{s1\lambda_l} \\ z_{s1\delta_l} \\ z_{s1ih_l} \\ z_{s2\lambda_l} \\ z_{s2\delta_l} \\ z_{s2h_l} \\ z_{s3\lambda_l} \\ z_{s3\delta_l} \\ z_{s3h_l} \\ z_{s4\lambda_l} \\ z_{s4\delta_l} \\ z_{s4h_l} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & (T_{s1}-t_0) & 0 & 0 \\ 0 & 1 & 0 & 0 & (T_{s1}-t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & (T_{s1}-t_0) \\ 1 & 0 & 0 & (T_{s2}-t_0) & 0 & 0 \\ 0 & 1 & 0 & 0 & (T_{s2}-t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & (T_{s2}-t_0) \\ 1 & 0 & 0 & (T_{s3}-t_0) & 0 & 0 \\ 0 & 1 & 0 & 0 & (T_{s3}-t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & (T_{s3}-t_0) \\ 1 & 0 & 0 & (T_{s4}-t_0) & 0 & 0 \\ 0 & 1 & 0 & 0 & (T_{s4}-t_0) & 0 \\ 0 & 0 & 1 & 0 & 0 & (T_{s4}-t_0) \end{bmatrix} * \begin{bmatrix} Z_{0_{\lambda_l}} \\ Z_{0_{\delta_l}} \\ Z_{0_{h_l}} \\ V_{0_{\lambda_l}} \\ V_{0_{\delta_l}} \\ V_{0_{h_l}} \end{bmatrix}$$

The cost of association now becomes

$$c_{i_1 i_2 \cdots i_S} = \sum_{s=1}^{S} [u(i_s) - 1]\ln(1 - P_{D_s}) + u(i_s)[(\mathbf{z}_{\mathbf{si}_s} - H(\hat{\mathbf{p}}_\mathbf{t}, \hat{\mathbf{v}}_\mathbf{t}, T_{si_s}))^T$$
$$\Sigma_s^{-1} (\mathbf{z}_{\mathbf{si}_s} - H(\hat{\mathbf{p}}_\mathbf{t}, \hat{\mathbf{v}}_\mathbf{t}, T_{si_s})) - \ln \frac{P_{D_s} \Psi_s}{2\pi |\Sigma_s|^{1/2}}],$$

where the normalized "residual" error squared term is given by

$$(\mathbf{z}_{\mathbf{si}_s} - H(\hat{\mathbf{p}}_\mathbf{t}, \hat{\mathbf{v}}_\mathbf{t}, T_{si_s}))^T \Sigma_s^{-1} (\mathbf{z}_{\mathbf{si}_s} - H(\hat{\mathbf{p}}_\mathbf{t}, \hat{\mathbf{v}}_\mathbf{t}, T_{si_s})),$$

and Σ_s^{-1} and $|\Sigma_s|^{1/2}$ provide the measurement error inverse covariances and the square root of the determinant of the measurement error covariance matrix, respectively.

Our goal is to find the most likely set of S-tuples such that each measurement is assigned to at most one target or declared false. Thus, we can recast this as the following generalized S-dimensional assignment problem

$$\min_{p_{i_1 i_2 \cdots i_S}} \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \cdots \sum_{i_S=0}^{n_S} c_{i_1 i_2 \cdots i_S} p_{i_1 i_2 \cdots i_S}$$

subject to:

$$\sum_{i_2=0}^{n_2} \cdots \sum_{i_S=0}^{n_S} p_{i_1 i_2 \cdots i_S} = 1; \text{ for } i_1 = 1, 2, \ldots, n_1$$

$$\sum_{i_1=0}^{n_1} \cdots \sum_{i_S=0}^{n_S} p_{i_1 i_2 \cdots i_S} = 1; \text{ for } i_2 = 1, 2, \ldots, n_2$$

$$\vdots \qquad \vdots \qquad \vdots$$
$$\vdots \qquad \vdots \qquad \vdots$$

$$\sum_{i_1=0}^{n_1} \cdots \sum_{i_{S-1}=0}^{n_{S-1}} p_{i_1 i_2 \cdots i_S} = 1; \text{ for } i_S = 1, 2, \ldots, n_S$$

where $p_{i_1 i_2 \cdots i_S}$ are binary event variables, such that $p_{i_1 i_2 \cdots i_S} = 1$ if the S-tuple $Z_{i_1 i_2 \cdots i_S}$ is associated to a target, otherwise, it is not considered, in the least-squares sense, to be associated to a target, thus, $p_{i_1 i_2 \cdots i_S} = 0$. As can be seen, the constraint set ensures that each scan measurement contributes to the cost calculation only once. Note that there are no constraints on the dummy measurements. Also note that through the use of dummy measurements, the association is now performed over sets of complete S-tuples. Thus, for a missed detection of a target in scan s, the corresponding S-tuple has $i_s = 0$.In the same vein, a false alarm is identified as an S-tuple with less than 3 (when S = 4) non-dummy measurements.

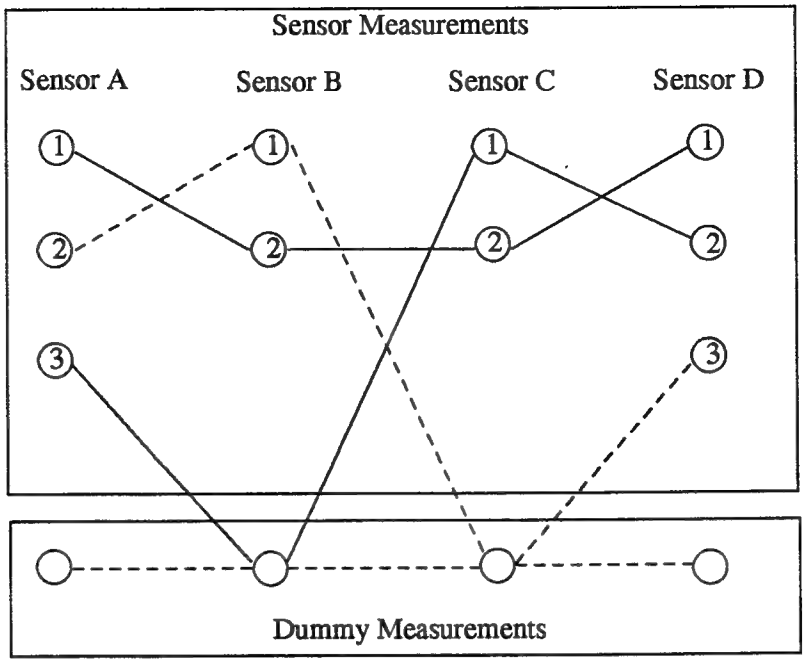


Figure 2. Interpretation of assignment.

As an example, in Figure 2 we illustrate the solution of a typical assignment problem. If we let the edges represent the actual assignment made by the S-dimensional ($S = 4$) assignment algorithm, and we note that the sensor scan lists under Sensors A..D could denote synchronous scans from 4 different sensors, or asynchronous scans from a single sensor, or some combination thereof, then the assignments can be interpreted as follows:

i) Assignment {1,2,2,1} represents a measurement 4-tuple from a target reported by all four sensors;

ii) Assignment {3,0,1,2} represents a 3-sensor detection. The target was missed by sensor B, but detected by sensors A, C, and D.

(iii) Assignments {2,1,0,0} and {0,0,0,3} both represent false alarms (spurious measurements).

6. Implementation

In this section we discuss some implementation details and initial results of the multiscan sliding-window S-dimensional tracker. For an overall block diagram of our implementation, refer back to Figure 1. The platform for this research was a Silicon Graphics 4D/35 Personal Iris MIPS R2000A/R3000 uniprocessor using the IRIX Release 4.0.5 System V operating system, and the language the algorithm was implemented in was C using ANSI's 3.10.1 C compiler.

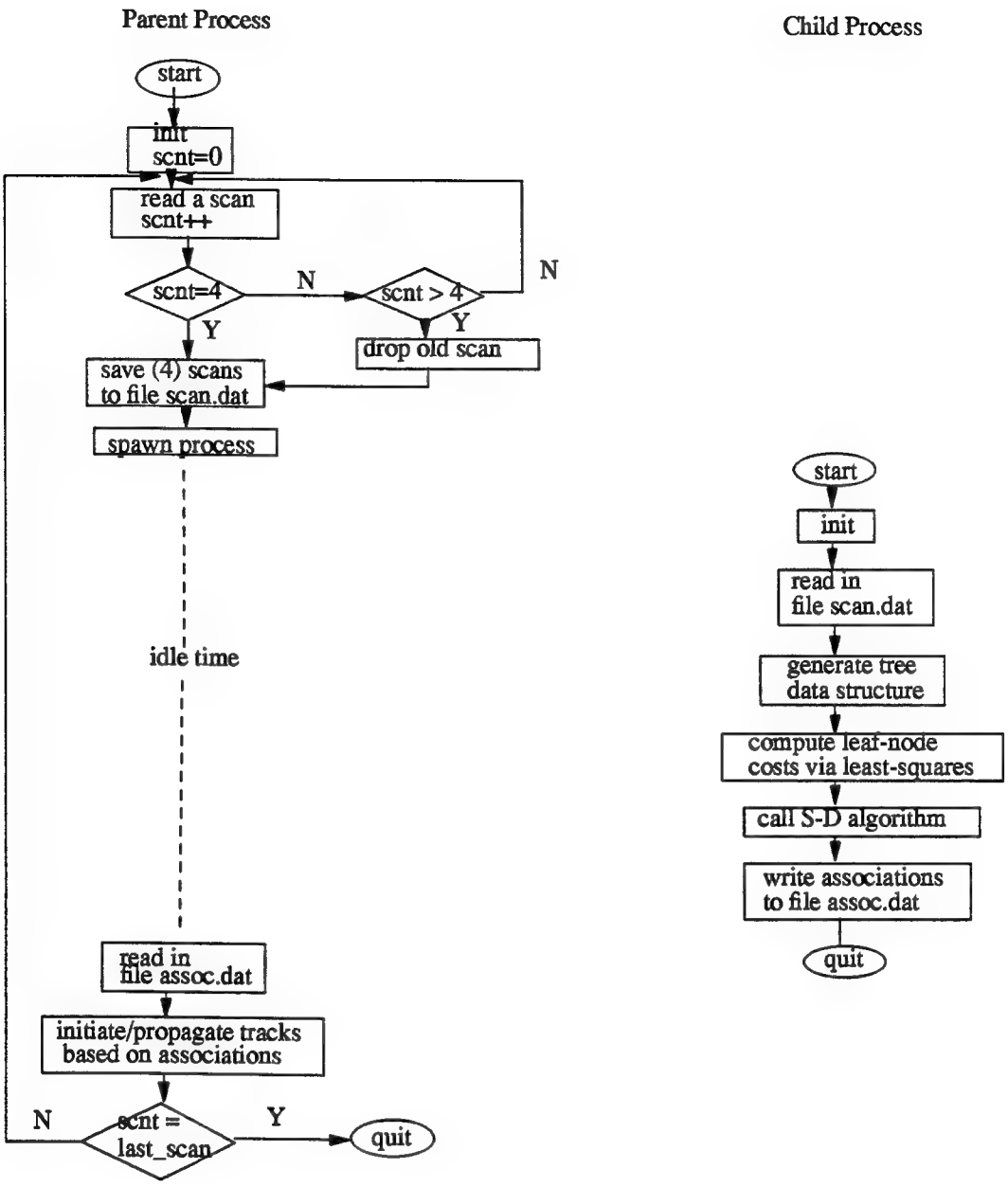


Figure 3. Flow chart of multiscan sliding-window tracker via the S-dimensional assignment algorithm.

Because of the incredible amount of recursion in the S-dimensional assignment algorithm, memory usage was at a premium. In order to survive the 24 Mbytes of RAM and the 150 Mbyte of swap space limitation on the system, we needed to partition the implementation of our multiscan sliding-window S-dimensional tracker into a multiprocess application, specifically, 2 processes (Figure 3). Initially, we invoke a process (say the parent) to iterate down the raw measurement data file, first obtaining all necessary extraneous parameters during the initialization phase of the implementation, and then grabbing 4 scans of

data at a time (in all but the initial case dropping out the oldest scan and obtaining the newest one). Once the four scans have been read, we dump this data to a file (i.e., scan.dat), spawn another process (say the child), where it now executes the S-dimensional assignment semantics. This in part includes the generation of the enormous tree data structure necessary as input into the S-dimensional assignment algorithm developed by Deb. Once the association is completed, the child process then dumps all candidate associations to a file (i.e., assoc.dat), and then dies, where upon its death, the parent resumes by first reading in the candidate associations, and then updating all pre-existing tracks and initializing on all new non-tracked associations. It then reads in another scan, drops out the old one, and continues with the aforementioned process until all scans are completed.

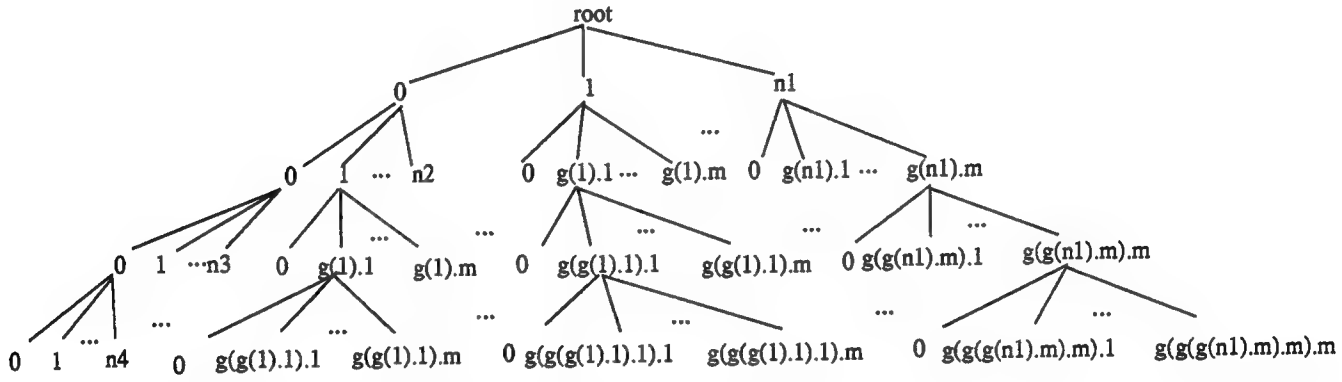


Figure 4. Input tree data structure to the S-dimensional assignment algorithm.

In the previous paragraph we made reference to the significant tree data structure necessary as input to the S-dimensional assignment algorithm. To give some sense of its significance, in Figure 4 we provide an general-type illustration of an example tree that would be generated. To interpret the tree, the level of the tree corresponds to the scan number within the window. For example, the *root* node is at level 0 and corresponds to "a fictitious" *0th* scan. Node *i* at level *k* in the tree corresponds to the *ith* measurement index of scan *k*. An edge connecting two nodes denotes the fact that the two measurements can potentially be associated together. For example, an edge interconnecting node *i* and node *j* at levels *k* and *k+1* corresponds to the case that the *ith* measurement from scan *k* can possibly be associated with the *jth* measurement from scan *k+1*. Two other notable things to mention: first, since every measurement from each of the scans can be associated with the dummy measurement, an outgoing branch originating from the measurement to the dummy of the next scan must exist; two, we use a global velocity gate around each scan measurement to reduce the number of measurements in the next scan that can potentially be associated with this measurement. For example, measurement 1 of scan 1 has m+1 measurements from scan 2 that it could possibly be associated with, specifically, the dummy measurement 0, plus $g(1).1, g(1).2, \ldots, g(1).m$ (note that $g(i).j$ denotes the *jth* measurement within the global velocity gate of a given scan's *ith* measurement). Notice that the "global velocity gate" for a given scan's dummy measurement includes all the

measurements of the next scan plus the dummy. This is necessary to handle all the possible assignments that may result from the S-dimensional assignment algorithm.

In our implementation of the multiscan sliding-window S-dimensional tracker, one particular aspect of the tracking process that may not be clear to the reader is how associated scan measurements are linked (or associated) to pre-existing tracks, if that were the case. Specifically, what is our tracker like? Conceptually, our tracking process via the S-dimensional assignment algorithm can be thought of as a series of track initiations (one per sliding window) followed by some sort of association-to-track linking (or association) semantics. In the case of say the *ith* sliding window, if the *jth* candidate association does not correspond to a pre-existing track established in at least one of the *i-1* sliding windows, then we simply initialize a new track based on this association. This simply means we assign to each measurement in the candidate association a new track id, and this id provides a source of input and a type of history that is utilized in future sliding windows when trying to link new associations to pre-existing tracks. So, for example, suppose Figure 2 provides a partial snapshot of the *ith* sliding window, and say the assignment {1,2,2,1} represents the *jth* candidate association not corresponding to a previous track. In this case, we would initialize a new track based on this association and assign to each measurement a track id number denoting this track initialization.

In the case that the *jth* candidate association corresponds to a target that has been previously tracked in the *i-1* sliding windows, then, conceptually, we need a sort of track-to-measurement_association linking (or association) process to handle the semantics of propagating a target along its track. In this situation, we have several cases. First I should note that if a measurement has never been assigned a track id, then it will have an initialized track id value of 0. Also, when trying to link the new scan measurement to a pre-existing track, only the first 3 scans track ids are inspected; the linking process really just trys to identify what to store as a track id for the new scan measurement, because after every sliding window, all candidate association measurements have a track id that has been updated. The general rule is as follows:

i) When inspecting the first three scan measurements of a candidate association within a particular sliding window, if at least two of the track ids are similar, then all measurements in the candidate association will have their track ids set to this value. For example, if partial assignment {1,2,2} from Figure 2 has at least two measurement track ids that are similar, say, for instance *k*, then measurements 1,2,2,1 from scans 1,2,3,4, respectively, will have their track id value set to *k*.

ii) If all three track ids are different, then they remain unchanged and the new scan measurement track id in the candidate association remains at 0 (i.e., untracked). So, from Figure 2, suppose track ids corresponding to the partial assignments {1,2,2} and {3,0,1} are [5,9,0] and [4,-,7], respectively. Then in each case, they remain unchange after this window finishes and the newest scan measurements 1 and 2 that correspond to the partial assignments {1,2,2} and {3,0,1} in the association have their track ids remain at 0. Note that in the second partial assignment there was no track id corresponding to the dummy measurement.

iii) and finally, if two of the three track ids are untracked (i.e., track id = 0), but the remaining track id is non-zero, then all four scan measurement track ids within this candidate association have their values set to this non-zero value. For example, if partial assignment {1,2,2} had corresponding track ids [7,0,0], then all four measurements of the association {1,2,2,1} would have their track id values set to 7.

In the first case, we simply employ a "majority rules" semantics. In the second case, we assume either the measurements are spurious, or that confusion exists in the score function and we do not want to continue to propagate this since we assume it is erroneous. In the third case, since we want to be optimistic, we allow such a track update to occur.

Given the data set provided by Rome Labs, because it was non-stressing, the results we obtained for the S-dimensional assignment algorithm with respect to track identification were excellent. Moreover, because of the built-in history within the association portion of the algorithm, we feel its effectiveness in more stressing data sets would still remain excellent. However, because of the computational explosion of the inherent multidimensional assignment problem in solving the S-dimensional assignment, the computational performance results we obtained here were not too good. In fact, in processing 55 scans of data (ranging from 22 to 43 measurements per scan) at 4 scans per sliding window, it took nearly 1.5 hours on the SGI platform previously mentioned. However, as reported by Deb, if the next S-dimensional process were to only build the additional dynamic tree data structure necessary, the number of computations per window would be one-tenth of what it currently is (i.e., $\approx$ 9 minutes processing). Specifically, the next S-dimensional process could occur without the need of recomputing a complete S-dimensional assignment problem (e.g., the solution to the next S-dimensional assignment problem can be initialized using the Lagrange multipliers heuristically obtained from the previous S-dimensional assignment problem). Furthermore, the inherent parallelism that can be exploited within the S-dimensional assignment algorithm is another source of improving its overall processing time. Thus, as a next logical step, the intended latter implementation and parallel development on high-performance (multiprocessor) computers, is the intended direction for this research. Because our results are only initial, we briefly provide the reader with results obtained by other researchers below.

In [14] a number of parametric studies are presented designed to explore the impact of window sizes, varying levels of measurement errors, probability of detection and probability of false alarms on solution quality and timings. For each of their studies, tracks were initiated using 4 to 8 scans and then extended using 3 to 7 scans, respectively. In track extension, the window was moved one scan at a time. For a 0.95 probability of detection, a 0.05 probability of false alarms, measurement errors in the interval [0.1,2.0], time interval between observations was every 10 seconds, with tracks initiated using 4 (5) scans extended using 3 (4) scans, the quality of track identification achieved using the 5/4 window (i.e., 5 scans on track initiation, 4 on track extension) was between 1% and 2% better than the 4/3 window, but this was offset by a threefold increase in the solution time. In that particular problem, having scan times every 10 seconds

apart, such a price can be paid without sacrificing real-time performance. For the 100 element target space, the solution quality and timing they demonstrated was reasonable. This is analogous to the results we obtained having a 4/4 window with similar parameter values (i.e., we considered a 0.99 probability of detection, and, as [14] notes, with a low probability of false alarms, decreasing the probability of detection does not substantially change the time necessary to solve the data association problem, although the quality of track identification may suffer due to lack of information). In another parametric study, having similar parametric values as above, they demonstrate the impact of the size of the track initiation and track extension windows. Their conclusion was that the advantages of a large track initiation window is removed after a few scans. Moreover, only marginal increases in performance were demonstrated for a 5-scan sliding window as opposed to a 3-scan and 4-scan window.

7. Conclusion

Future work would encompass several modifications to the current implementation of the S-dimensional assignment algorithm developed by [4]. Three primary improvements would be the intended heuristic implementation of the next S-dimensional assignment from the previous S-dimensional assignment, an alternative tree representation and various tree pruning strategies (i.e., the dummy branches) to the extremely large tree created in the current version, and a parallel implementation of the algorithm. In the latter case, because the general multidimensional assignment problem is an inherent problem of the S-dimensional assignment algorithm, much of the potential parallelization of the general problem also applies to the S-dimensional problem. For the second problem, one observation made, in particular, would be to significantly reduce the tree representation in the current version; an alternative strategy in representing the dummy nodes in each scan would greatly improve the performance of the S-dimensional algorithm. Instead of creating the dummy node branches within the tree for all measurements in each of the scans, store with each measurement a special structure denoting special case semantics for the dummy nodes. This would eliminate a large number of hopeless branches to pursue during the relaxation and cost computation of the S-dimensional algorithm, thus eliminating many of the unnecessary recursive calls.

Acknowledgments

I would sincerely like to thank Dr. Yakov Bar-Shalom, Dr. Krishna Pattipatti, Somnath Deb, YRK, Vince Vannicola, Rick Gassner, Mark Barnell, and, in particular, Gerald Bright for all the incitefullness, ingenuity, and wisdom given during my brief summer tenureship at Rome Laboratory.

References

[1] Bar-Shalom, Y., ed., *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Dedham, MA., 1990.

[2] Bar-Shalom, Y., ed., *Multitarget-Multisensor Tracking: Applications and Advances*, Artech House, Dedham, MA., 1992.

[3] Bar-Shalom, Y., and T.E. Fortmann, *Tracking and Data Association*, Mathematics in Science and Engineering, vol. 179, Harcourt Brace Jovanovich, Boston, MA. 1988.

[4] Deb, S., K.R. Pattipatti, and Y. Bar-Shalom, "A S-dimensional assignment algorithm for track initiation," *Proceedings of the IEEE Systems Conference*, Kobe, Japan, September 1992.

[5] Deb, S., K.R. Pattipatti, and Y. Bar-Shalom, "A new algorithm for the generalized multidimensional assignment problem," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Chicago, 1992.

[6] Deb, S., K.R. Pattipatti, Y. Bar-Shalom, and R.B. Washburn, "Assignment algorithms for the passive sensor data association problem," *Proceedings of the 1989 SPIE Conference*, Orlando, FL., March 1989, pp. 231 - 243.

[7] Gilbert, K.C., and R.B. Hofstra, "Multidimensional assignment problems," *Decision Sciences*, vol. 19, 1988, pp. 306 - 321.

[8] Morefield, C.L., "Application of 0-1 integer programming to multitarget tracking problems," *IEEE Transactions of Automatic Control*, vol. AC-22, no. 3, June 1977, pp. 302 - 312.

[9] Pattipatti, K.R., S. Deb, Y. Bar-Shalom, and R.B. Washburn, "A new relaxation algorithm and passive sensor data association," *IEEE Transactions on Automatic Control*, vol. 37, no. 2, February 1992, pp. 198 - 213.

[10] Pattipatti, K.R., S. Deb, Y. Bar-Shalom, and R.B. Washburn, "Passive sensor data association using a new relaxation algorithm," *Multitarget-Multisensor Tracking: Advances and Applications*, Y. Bar-Shalom, ed., Artech House, Dedham, MA., 1992.

[11] Poore, A.B., "Data association problems in multisensor data fusion and multitarget tracking," *AFOSR Summer Research Program Final Report*, Rome Lab, August 1992, pp. 89.1 - 89.20.

[12] Poore, A.B., and N. Rijavec, "Multitarget tracking, multidimensional assignment problems, and lagrangian relaxation" *Proceedings of the 1991 SDI Panels on Tracking*, 1991, pp. 3.51 - 3.74.

[13] Poore, A.B., and N. Rijavec, "Multi-target tracking and multi-dimensional assignment problems," *Proceedings of the 1991 SPIE Conference on Signal and Data Processing of Small Targets*, vol. 1481, 1991, pp. 345 - 356.

[14] Poore, A.B., N. Rijavec, and T.N. Barker, "Data association for track initiation and extension using multiscan windows," *Proceedings of the 1992 SPIE Conference on Signal and Data Processing of Small Targets*, 1992.

[15] Poore, A.B., N. Rijavec, M. Liggins, and V.C. Vannicola, "Data association problems posed as multidimensional assignment problems: problem formulation," *Proceedings of the 1993 SPIE Conference on Signal Processing, Sensor Fusion, and Target Recognition II*, Bellingham, WA., 1993.

[16] Sittler, R.W., "An optimal data association problem in surveillance theory," *IEEE Transactions on Military Electronics*, vol. AES-11, no. 1, January 1981, pp. 122 - 130.

[17] Yeddanapudi, M., Y. Bar-Shalom, K.R. Pattipatti, T. Kirubarajan, and S. Deb, "MATSurv: Multisensor Air Traffic Surveillance," *Rome Lab Progress Report*, April 26, 1994, pp. 1 - 36.

Mr. Steven J. Pratt
Report
Unavailable

Francis X. Reichmeyer
Report
Unavailable

SELF-SUSTAINED PULSATION AND HIGH SPEED OPTICAL NETWORK NODE

David H. Sackett
Masters Graduate Student
Department of Electrical Engineering

Rochester Institute of Technology
One Lomb Memorial Drive
Rochester, NY 14623

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

September 1994

SELF-SUSTAINED PULSATION AND HIGH SPEED OPTICAL NETWORK NODE

David H. Sackett
Masters Graduate Student
Department of Electrical Engineering
Rochester Institute of Technology

ABSTRACT

During my research period at Rome Labs I had worked on three projects simultaneously. One of the projects studied Self-Sustained Pulsation(SSP) from laser diodes. SSP was studied as a means of creating an inexpensive high bandwith optical communication system. The other two projects were related high speed optical networks. Two components of the electrical control system were evaluated and modified to operate in a network running at 2.5 GHz bit rate.

Self-Sustained Pulsation(SSP) is an intrinsic intensity modulation occurring in laser diodes when biased above the required lasing threshold current. SSP is generally considered undesirable in continuos-wave(CW) laser diodes. The phenomenon was found at several investigated wavelengths. SSP can be used as a carrier wave for high bandwidth modulated information transfer. During the work at Rome Labs a microwave feedback system was installed to enhance the SSP. The SSP carrier was then used to successfully transmit and receive AM and FM test signals.

Previously demonstrated at Rome Labs was a two port network node which passed optical data without the need for regeneration. This system operated at a 1.2 GHz transmission speed and will soon be operating at 2.5 GHz. Several of the circuits were in need of optimization so as to perform satisfactorily at the 2.5 GHz bit rate. Three electro-optic switch drivers were in need of such a review. These analog circuits received a small input signal ($0 \rightarrow 1$ Volt) and were required to switch the output from $1 \rightarrow 9$ Volts in approximately 10 mseconds. The original circuit proved to be sufficient for this task. The only changes required were in the board layout and a minor change in rated power supplies.

The AC coupler, also to be part of the 2.5 GHz optical communications node, was needed to provide drive current and isolation. The original AC coupler circuit provided isolation from a kilohertz range through 100 MHz. The completed circuit successfully coupled data with reasonable accuracy within the range of 50 MHz through 3 GHz bit rates. At the target speed of 2.5 Ghz the device demonstrated an acceptable eye-diagram.

INTRODUCTION

I.

Most laser diodes of any wavelength produce an intensity modulation in the 2 to 7 GHz range, Self-Sustained Pulsation(SSP). This pulsation is often ignored or even suppressed in most laser diode applications. The modulation can, however, be used as an optical carrier for a low cost high bandwidth optical communication network. When the SSP is predictably controlled a simple AC bias tee allows an electrical signal to be placed on a high bandwidth optical carrier with NO additional hardware. The information may be transmitted directly with AM or using FM modulation. The transmitter and receiver technology is simply a microwave version of common radio.

The open loop system consisted of a laser diode being driven by an adjustable current source; temperature control is also desirable. The information to be transmitted is simply added to the DC supply with a bias tee. A high speed optical detector (with fiber optic input) translated the intensity signal into a microwave electrical signal. The signal was then amplified, demodulated and passed to a spectrum analyzer. The relationship between the laser modulation intensity and the information amplitude is actually a square relationship. Therefore the detection system requires slight modification from a standard AM or FM receiver. This experiment utilized a simple differentiator to recover the signal with recognizable results.

There are some challenges is this method of communication. The linewidth of the open loop SSP is on the order of 500 MHz, and the modulated power is only a fraction of the CW power output of the laser. To reduce these two problems a feedback loop may be added to the system. The feedback system used was electrical. The signal from the high speed optical detector was amplified again and added to the DC bias current along with the information signal. The feedback was successful in increasing the amplitude by several orders of magnitude and in reducing the linewidth to 20 KHz. This created very distinct SSP and clear modulation with discernible demodulation.

The feedback loop was found to affect the SSP frequency range. Excessive feedback would fix the SSP frequency independent of DC bias. Filtering of the harmonics in the feedback system added stability and control of the SSP frequency. Several weeks were devoted to the development of low-pass and band-pass microwave filters for the feedback loop.[1]

The exact range of SSP bias current required and the frequency of SSP varies unpredictably for each laser diode. The first stage of the investigation involved 'mapping' many laser diodes of different wavelengths. Not all DC current bias levels produce SSP. And self-pulsation, once obtained, was often stable for limited frequency ranges. The range of SSP frequencies obtainable varied with every laser diode. In addition some diodes produced a smooth range of SSP frequencies while others displayed a jumping of frequencies. 780 nm and 860 nm and 1300 nm laser diodes where investigated. SSP was found at all of these light frequencies.

II.

Rome Labs has previously demonstrated a node structure for an optical communication network operating at 1.2 GHz bit rate. The node circuitry converted only the pertinent header information to electricity. This information is needed to set the electro-optic switches. Optical header and data are then passed through the node without the need for regeneration. The same node structure will be demonstrated operating at a 2.5 Ghz bit rate. Several components required modification to operate at twice the previous speed.

The driver circuit consists of a passive pull-up NMOS inverter followed by a discrete CMOS inverter. The circuit drove a $1 \rightarrow 9$ Volt electro-optic switch from a one volt supply. The circuit performed the switch in under 10 mseconds. The first stage of this project involved gaining an understanding of the limitations of the circuit and evaluating its ability to operate fast enough for the 2.5 GHz network. The circuit was found to operate with sufficient speed. Only minor modifications were made. The printed circuit board(PCB) was optimized by removing an extraneous power supply, lowering power supply levels and power dissipation, using smaller heat sinks, and re-arranging the PCB traces for a more intuitive layout. The fundamental circuit operation was not altered. The new circuit was fabricated.

III.

Another circuit for the 2.5 GHz optical network node was an AC coupler circuit. The main component of the circuit was a surface mount ECL device surrounded by passive surface mount components. The AC coupler was needed to provide drive current and isolation between two segments of the electrical control system for the gate switches. The original AC coupler circuit provided isolation from a kilohertz range through 100 MHz.

The first stage of investigation involved a careful experimental and theoretical analysis of the original circuit. Only passive components could be modified. Three possible problems were identified. First, the desired high frequency demanded low inductance in the circuit. The larger capacitors were thought to be contributing to extra inductance. Secondly the coupling (series) capacitance was higher than the chip specifications. Lastly, there was found a current flow imbalance caused by the inverted logic output not being used.

Three minor changes were made. The inverted output logic was provided with a dummy load to balance current switching between the logic output and the inverted logic output of the ECL device. The second modification was the changing of certain capacitor values and the addition of small capacitors in parallel to over come inductive delays in the larger capacitors. The final modification was to increase the paths to ground and insure the security of distributed grounding.

The completed circuit successfully coupled data with reasonable accuracy within the range of 50 MHz through 3 GHz bit rates. At the target speed of 2.5 Ghz the device demonstrated an acceptable eye-diagram.

CONCLUSION

I.

Self-Sustained Pulsation was found to exist in laser diodes at 780, 860, and 1300 nm. With an electrical feedback system linewidths of 20 KHz were obtained for an optical carrier in the 2 GHz to 7 GHz range. AM and FM signals were successfully modulated by the optical carrier and demodulated at the receiver. This system has shown the potential of a very inexpensive high bandwidth intensity modulation scheme. Further investigation is required to increase the efficiency of the SSP and to understand the feedback variables more quantitatively. In addition a through analysis of signal to noise ratio is required.

II.

The modifications performed to the electro-optic switch driver circuit were mostly aesthetic. The only significant modification to the original circuit was to reduce the power dissipation in the regulation system.

III.

The completed circuit demonstrated AC coupling through 3 GHz. The input to the circuit is one $Volt_{p-p}$ with any DC offset. The circuit drives an output of three Volts. The circuit is presently propagating very close to the ECL chip's rated speed. In order to generate faster response a new technology level will be required.

REFERENCES:

[1] Rome Laboratory Internal Report; David H Sackett

ACKNOWLEDGMENTS:

I would like to acknowledge Dr. Guifang Li, Dr. Ray Boncek, and Andew Pirich for their support and the opportunity to pursue the three projects detailed in this report.

A STUDY OF HIGH SPEED POLARIZATION ROTATORS FOR USE IN AN OPTICAL INTERCONNECTION NETWORK

Paul E. Shames
Department of Electrical and Computer Engineering

University of California, San Diego
La Jolla, CA 92093

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

August 1994

A STUDY OF HIGH SPEED POLARIZATION ROTATORS FOR USE IN AN OPTICAL INTERCONNECTION NETWORK

Paul E. Shames
Department of Electrical and Computer Engineering
University of California, San Diego

Abstract

The polarization rotation characteristics of ferroelectric liquid crystals (FLC) and lanthanum-modified lead zirconate titanate (PLZT) ceramic are studied for use in an optical interconnection network. A single pixel FLC device is found to have greater than 1000:1 contrast ratio with a rise time of 33 μs running at 7 kHz. A 10×10 FLC array, illuminated by IR light, has better than 100:1 contrast and 135 μs response time at 67 Hz. PLZT optical shutters have been reported to have response times of 1 μs and contrast ratios of 1500:1.

A STUDY OF HIGH SPEED POLARIZATION ROTATORS FOR USE IN AN OPTICAL INTERCONNECTION NETWORK

Paul E. Shames

Introduction

As the size of communication systems and computer memory grow, the use of optics as a means of data transmission and interconnection becomes more attractive [1]. At UCSD we are developing an optical interconnection network that uses a series of polarization selective switches [2]. These interconnects, made of birefringent computer generated holograms (BCGH), direct light depending on the state of the polarization of the incoming beam. This is controlled by a spatial light modulator (SLM) performing polarization rotation. The two types of SLMs that are looked at here are FLC, with reported response times of 70 μs [3] and PLZT with a response time of <1 μs [4,5].

In what follows, the BCGH switch is described and two polarizers and FLC devices are analyzed. The polarization characteristics of PLZT are reviewed and further testing is proposed.

Selective Polarization Device and Network

The design for a polarization selective 2×2 switch is shown in Figure 1. One entering beam has vertical polarization and the other has horizontal. The first BCGH combines and focuses the beams into the SLM, which either rotates the polarizations or not. The second BCGH directs the outputs to their destinations. The speed of the switch is limited by the response time of the polarization rotator, which is the only active component.

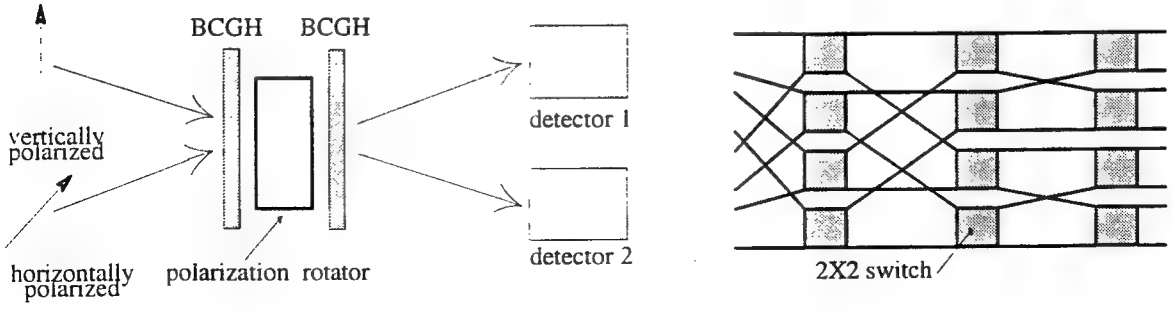


Figure 1: On the left is a 2×2 Optical Switch that uses two BCGHs and a polarization rotator to direct the polarized light to detectors. On the right is an 8×8 Benes network that uses 2×2 switches in series.

These switches will be used in a network that has many components in series (see Figure 1). The Benes network shown is one of many possible configurations that can take advantage of the 2×2 Optical Switch. High transmittance is required so that the power of the signal does not diminish too drastically after propagating through the network. When a vertically polarized beam enters the device it is directed to one of two destinations. If part of that beam becomes depolarized then some of it will be sent to the opposite destination, i.e. noise. In order to maintain a low signal to noise ratio the rotators must have high contrast ratios.

Polarizers

The beams need to be linearly polarized before they enter the switch, so we begin by examining two different types of polarizers. Separate testing of the birefringent computer generated hologram (BCGH) was done using 515 nm light, but several other wavelengths are of interest, particularly 1300 nm, which is useful in fiber-optics. A dichroic sheet polarizer absorbs one polarization and allows the other to pass through. A Melles-Griot device, using a 633nm HeNe laser, had a measured extinction ratio of 3.6×10^{-5} (max = 830 μW and min = 0.03 μW). Using an Argon Ion laser with a line filter centered at 515 nm (10 nm FWHM) the dichroic polarizers had a 2.1×10^{-4} extinction ratio. Glan-Thompson (GT) polarizing prisms, which deflect one polarization and allow the other to propagate, had an extinction ratio of 2.6×10^{-6} for 633 nm, 3.7×10^{-6} at 515 nm and 2.2×10^{-4} using 1319 nm from a diode pumped solid state ring laser. The dichroic polarizers do not absorb well in the IR range. Near-IR polarizers, boasting extinction ratios of $<10^{-2}$, made of a metallic thin film or wire grid are available from Melles-Griot. The dichroic polarizers have a 33% transmission for unpolarized light and the GT transmit 36%. Although the GT polarizers have better contrast, the dichroic polarizers are far more compact and have a wider useful field angle.

FLC Basics

Liquid crystal displays generally are of the twisted nematic variety, where the light follows the twisting of the crystal as it propagates through the material. An applied voltage changes the twist, thus altering the polarization rotation characteristics of the crystal. A ferroelectric liquid crystal is of the smectic variety, where the crystals line up in layers of similar orientation (under no applied field crystals in different layers may have different orientations). The permanent electric dipole of ferroelectrics means that the electric dipole of each molecule will line up in the direction of an applied field, aligning all of the crystals. By switching an electric field from positive to negative the crystal alignment can be reoriented very fast. To get polarization rotation one would choose one polarity to be aligned with an optical axis of the crystals and the other to be at 45°, giving the maximum rotation. The FLC thickness is chosen according to the amount of rotation desired and the wavelength being used.

Single Pixel

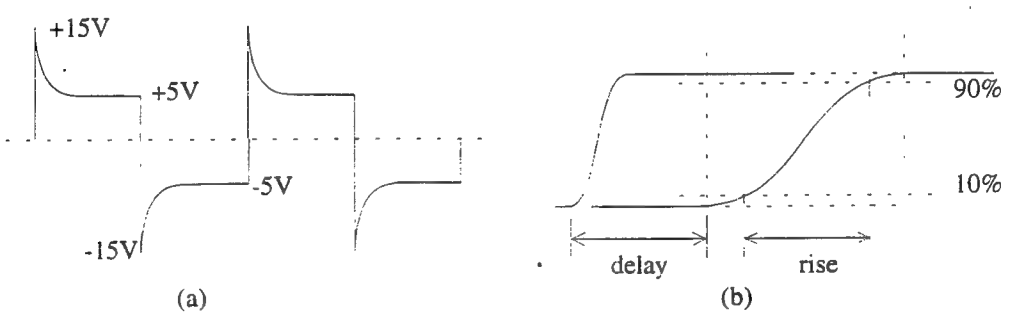


Figure 2: (a) DT DR50 drive wave form and (b) delay and rise time diagram

The DisplayTech (DT) LC050AC Light Valve is a single pixel FLC device with attached adjustable dichroic polarizers. The contrast ratio between the 'on' and 'off' states was measured at 6153:1 using 515 nm light. This figure was the maximum attained, typical contrast ratios were around 1000:1. The operating wavelength range of the device is 400 nm to 700 nm, with an optimum at 520 nm. The device comes with it's own driver that uses a ±15 V pulse which exponentially decays (in 300 μs) to ±5 V, as seen in Figure 2a. While 15 V is needed to optimize the switching, only 5 V is needed to hold the FLC alignment

and maintaining the lower voltage increases the life of the FLC. For comparison, we also drove the FLC with a square wave at various voltages (±5 V, ±10 V and ±15 V).

Of primary interest in our application is the response time and the contrast ratio of the FLC. For response time, the 10-90% rise time is used (see Figure 2b). The fall times were similar to the rise times. Figure 3 shows that as the frequency of the driving signal increases the response time of the FLC decreases. However, looking at (a) through (d) of the figure, it's seen that at high frequency the FLC is switched to the 'off' state before it has completely reached the 'on' state, and vice a versa.

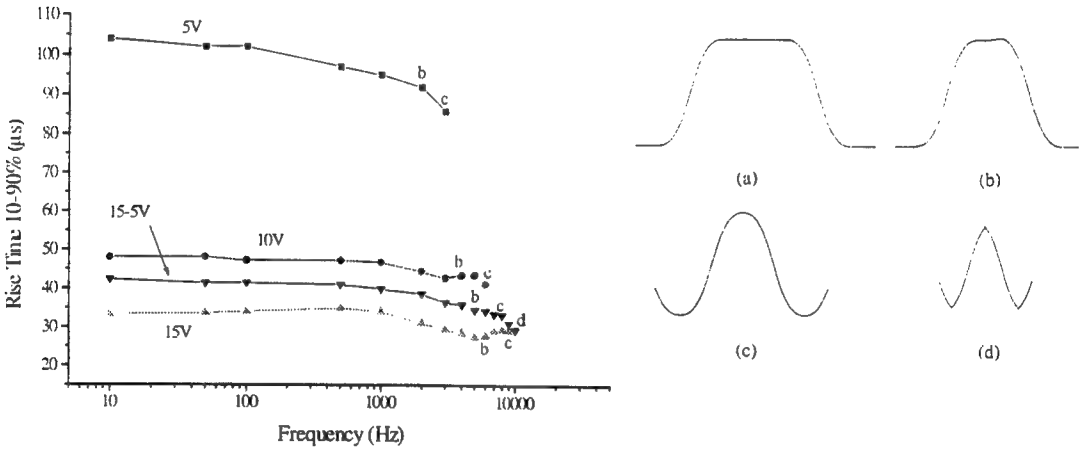


Figure 3: The 15-5V points represent the ±15-5 V waveform from the DR50 driver. The 5V, 10V and 15V points represent the ±5, ±10 and ±15 V square wave points. As the frequency increases, the wave form of the optical output changes from a relatively square wave (a) to a triangular wave (d).

It is difficult to measure power at high frequency because of the slow time averaging of a power meter. The photodiode, used to measure response times, creates a current that is displayed on the oscilloscope as a voltage. The difference in voltage (ΔV) between the 'on' and 'off' states is proportional to the optical power contrast. Figure 4 shows that the contrast (or ΔV) increases by approximately 10% before decreasing rapidly at higher frequencies. Looking at the maximum contrast points we see that the ±15 V square wave gives the shortest response time of 27.7 µs at 6 kHz, however this voltage is too high for prolonged use. The DR50 ±15-5 V waveform has a maximum contrast at 7 kHz and a response time of 33.3 µs. DisplayTech specifications for the LC050AC give typical response and delay times of 35µs and a contrast ratio of 1000:1 (for white light). The ±5 V signal, at 2 kHz, has a 92 µs response at maximum

contrast. Periodic measurements of the delay time between the beginning of the drive wave and the beginning of the response curve consistently are <50 μs.

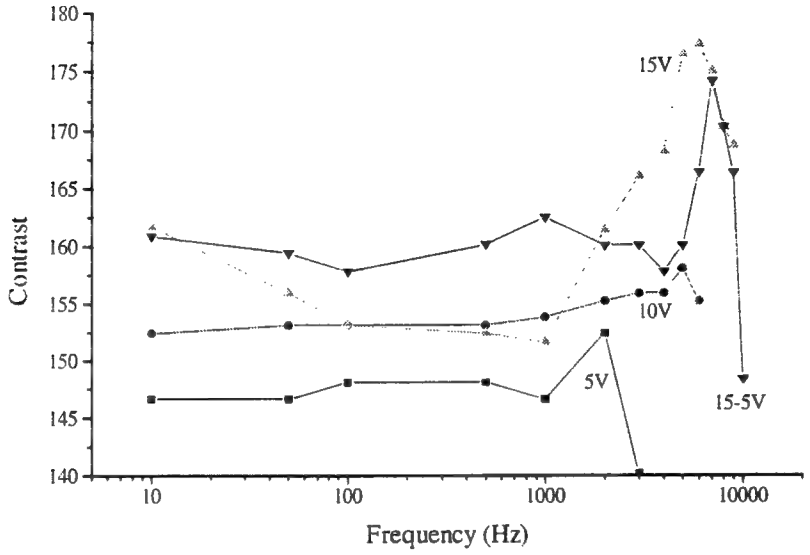


Figure 4: Contrast (voltage difference) between the 'on' and 'off' states increases until 6-7 kHz for 15V and 15-5V drivers and then drops very quickly off the scale.

Not shown on Figures 3 and 4 are the points at 20 kHz. While the 10-90% rise time was 18.6 μs, the contrast was reduced by 70% (off the scale of Figure 4). The waveform here is similar to that shown in Figure 3d. The output signal slope peaks and changes to negative before the device has turned completely 'on'. The FLC oscillates between being partially 'on' and partially 'off'. Consequently, the rise time is considerably shortened and the contrast ratio is greatly reduced.

Also of interest is the transmissivity of the FLC. Using a beam of 4.25 mW from the Argon Ion laser, and with the polarizers removed from the device, the output is 3.60 mW, which is 85% transmission. Reflection from the FLC is 0.40 mW. With an AR coating applied, 94% transmission should be possible.

10×10 Array

A 10×10 DisplayTech FLC array was also examined. The device is designed to work in the 1200 to 1400 nm range. In order to illuminate many pixels at the same time, the collimated 1319 nm laser beam was expanded by a factor of 10:1. To counter diffraction effects (particularly when looking at the cross-talk between pixels) the array was imaged onto the detector surface (see Figure 5). An optical power meter was used to measure contrast, a photodiode for response times and a ccd camera was used to look at cross-talk between pixels.

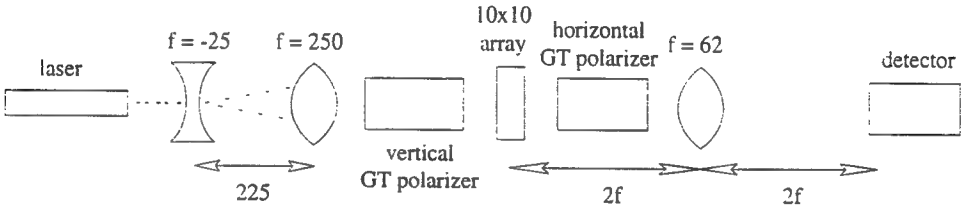


Figure 5: Experimental setup for 10×10 FLC array. The photodiode was placed at the focal point of the imaging lens. Units are in mm.

Maximum contrast, when looking at a beam more than 30 square pixels in size, was 124:1. The device is designed for the incident light to be vertically polarized. Looking at Figure 6 we see the contrast drop off sharply as the polarization of the incident light is varied. At ±5° the contrast is approximately 20:1. The slight increase of contrast at higher angle is due to addition of the horizontal component. For vertically polarized light that is 20° from normal incidence, the array maintains a relatively good contrast of 73:1. Due to the width of our incident beam we can not look at higher angles without getting beam blockage from the edges of the device.

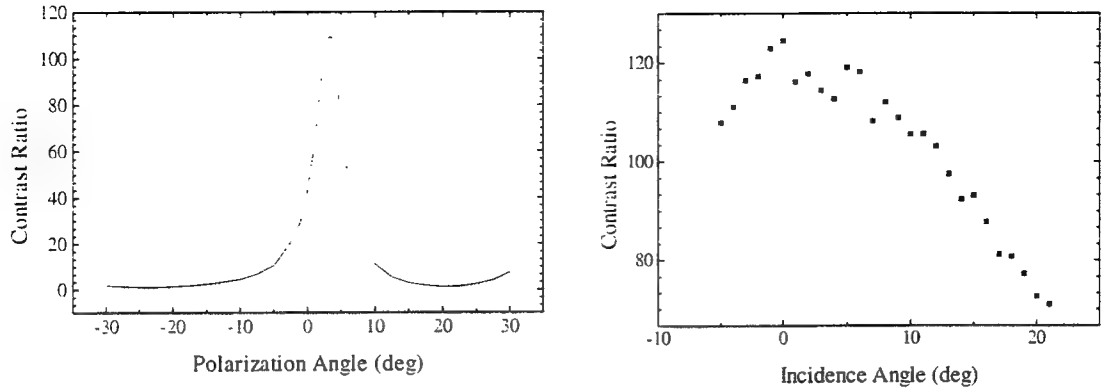


Figure 6: On the left, the slight displacement from center (approx. 3°) of the maximum contrast either represents the true vertical position (vertical was only approximately established) or else a slight misalignment of the FLC array. The right hand plot shows >70:1 contrast at 20° from normal.

The array comes with it's own driver. It's smallest time step is 15 ms which gives a driving rate of 67 Hz. At this speed the response time of the array is 135.7 μs. Random testing of individual pixels gave an average response time of 135.9 μs. the standard deviation being 1.3 μs (1%), and the largest deviation being only 3%.

Another characteristic of interest is the amount of cross-talk between pixels. This can come about because the electric field across one pixel can 'leak' over to surrounding pixels. However, it is obvious from the image through a ccd camera (see Figure 7) that any cross-talk is negligible. The slightly higher intensity seen around the edges of the pixels is probably due to diffraction.

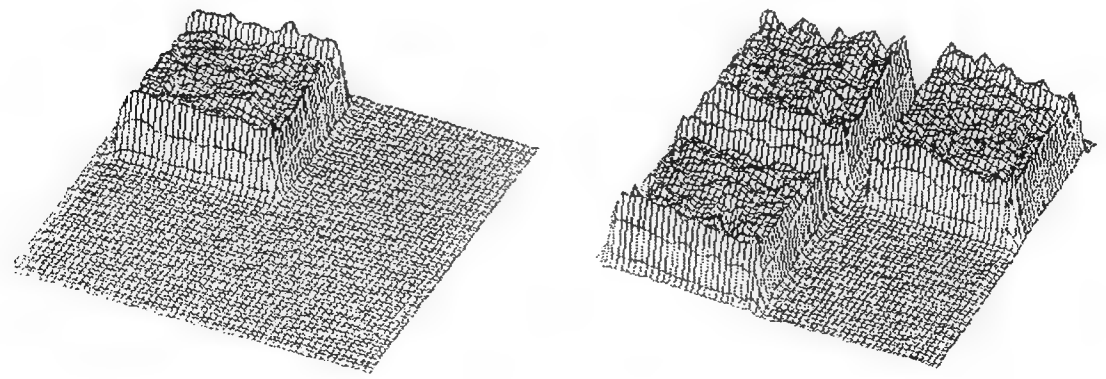


Figure 7: Relatively no cross-talk between pixels can be seen in the ccd camera image of four pixels. In the left image one pixel is 'on' and three are 'off'. in the right image three pixels are turned 'on'.

PLZT

PLZT ceramic has field-induced birefringence [6]. When no field is applied across the material the index of refraction is isotropic. Under an applied field, the index along the crystal axis perpendicular to the field increases while the index that is parallel remains unchanged. Vertically polarized light that is incident at 45° has it's polarization rotated after passing through the PLZT that has an applied field. When no field is present the polarization is unchanged. The standard method for applying an electric field is by using either metallic or ITO electrodes, in an interdigital pattern (see Figure 8a), on the optical surface of the PLZT.

The Motorola Solid Ceramic Shutter 1108A, made of PLZT with bimetallic electrodes in an interdigital pattern on both optical faces, boasts a response time of 1 μs and a contrast ratio of 1500:1 at a half-wave voltage of 400 V [7]. Kunihiro Nagata, at the National Defense Academy of Japan, used copper electrodes produced by a combination of photo-lithography and electroless plating on 9/65/35 PLZT. He measured <2 μs response times at a half-wave voltage of 190 V [4]. This method of applying the electric field can result in an uneven variation of the index between the electrodes.

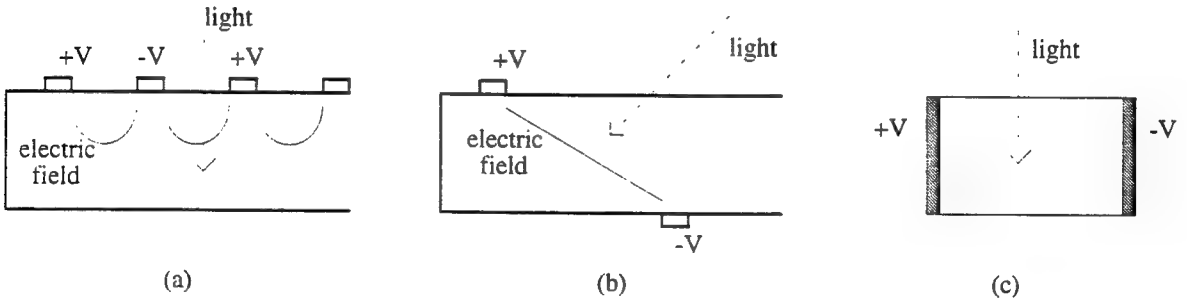


Figure 8: (a) Interdigital electrode pattern on one optical surface, (b) electrodes on opposite surfaces at 45° and (c) electrodes on sides of PLZT.

A possible solution is a digital electrode pattern that places the positive and negative elements on opposite sides of the PLZT wafer (Figure 8b). The electric field would be at 45° to the surface of the wafer and if light was incident at 45° the maximum rotational properties of the PLZT could be used. Another method of applying the field is to place the electrodes on the sides of the PLZT (Figure 8c).

Conclusion

The FLC polarization rotator can provide contrast of 1000:1 at switching speeds of up to 7 kHz while being driven by a modest ±15V supply. While somewhat sensitive to polarization alignment, the devices provide good contrast at angles up to 20° from normal. This provides some flexibility in design and packaging of our polarization selective BCGH switch. The 94% transmissivity of the FLC is within the required range for a network.

PLZT promises to be an even faster material for polarization rotation. Commercial devices already boast response times of 1 μs. There was not sufficient time in the course of this experiment to analyze PLZT directly. In the hope of getting better contrast and lower half-wave voltages, Rebecca Bussjager, at the Photonics Center at Rome Laboratory, is fabricating and testing the 45° opposite side placement of the electrodes. At UCSD we plan on experimenting with the side placement of the electrodes.

Acknowledgments

The author would like to thank Rebecca Bussjager , Joseph Osman and Richard Michalak and others at the Photonics Center and Albert Jamberdino and the fine people at Intelligence and Reconnaissance at Rome Laboratory for their kind assistance and invaluable advice. Thanks also goes to DisplayTech for the use of their LC050AC optical shutter.

References

1. M. Feldman, S. Esener, C. Guest, and S. H. Lee. "A Comparison of electrical and free space optical interconnects." *Applied Optics* **27**, 1742-51 (1988).

2. F. Xu, J. Ford and Y. Fainman. "Polarization selective computer generated holograms and applications," SPIE International Symposium on Optics, Imaging and Instrumentation, Paper 1992-19, San Diego, CA (July 1993).

3. M.G. Roe and K.L. Schehrer. "High-speed and high-contrast operation of ferroelectric liquid crystal optically addressed spatial light modulators," *Optical Engineering* **32**(7), 1662-1667 (July 1993).

4. K. Nagata. "PLZT Optical Shutter with copper electrodes by electroless plating," *Ferroelectrics* **109**, 247-52 (1990).

5. J. Phillips. "Advanced technology: PLZT ceramics." *Information Display*, Motorola Ceramic Products (April 1989).

6. G. Heartling. "Piezoelectric and Electrooptic Ceramics," in *Ceramic Materials for Electronics*, R. Buchanan (ed.), New York, Marcel Deker (1986).

7. Motorola Ceramic Products SCS1108A Performance Specifications (1988).

A PROGRAM PLAN FOR TRANSMITTING HIGH-DATA-RATE ATM/SONET SIGNALS OVER THE ACTS

Valentine Aalo
Assistant Professor
Electrical Engineering Department
Florida Atlantic University
500 North West 20th Street
Boca Raton, FL 33431

Okechukwu Ugweje
Electrical Engineering Department
Florida Atlantic University
500 North West 20th Street
Boca Raton, FL 33431

Final Report For:
Summer Research Program
Rome Laboratory

Sponsored By:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

U.S. Air Force Rome Laboratory

September 1994

A PROGRAM PLAN FOR TRANSMITTING HIGH-DATA-RATE ATM/SONET SIGNALS OVER THE ACTS

Valentine Aalo
Assistant Professor
Electrical Engineering Department
Florida Atlantic University

Okechukwu Ugweje
Electrical Engineering Department
Florida Atlantic University

Abstract

The feasibility, desirability and usefulness of Asynchronous Transfer Mode and Synchronous Optical Network transmission protocols over the Advanced Communications Technology Satellite (ACTS) was studied. A program plan for the transmission of Asynchronous Transmission Mode and Synchronous Optical Network signals at high data rates via the ACTS satellite was developed for the U. S. Air Force Rome Laboratory. The high data rate terminals will transmit and receive signals at DS-3 (45 Mbps) and OC-3 (155 Mbps) over the NASA's ACTS.

A PROGRAM PLAN FOR TRANSMITTING HIGH-DATA-RATE ATM/SONET SIGNALS OVER THE ACTS

Valentine Aalo
Okechukwu Ugweje

1 INTRODUCTION

Communication systems have developed at a very rapid pace over the last decade. One of the outcomes of this development is in the area of broadband packet switching. In the last few years, research activities in this field have been very active. Hardly any publication presently on switching networks can be opened without the mention of ATM. ATM is now regarded as the communication protocol of the future and all efforts are being made by researchers to find more effective use of this new technology.

The objective of this program plan is to study the feasibility, desirability and usefulness of using ATM/SONET transmission protocols over the NASA's Advanced Communication Transmission Satellite (ACTS) for multimedia application at HDR. This objective is a subset of a proposed high speed network under development which will link the assets of the Air Force at Rome Laboratory (RL), the Army at Communications Electronic Command (CECOM), and the Navy at Naval Research and Development (NRaD). This network has the potential of becoming a world wide broadband network connecting a large number of remote tactical sites (i.e. U.S. Navy ships) with voice, video and data. This objective will then be integrated into the Technical Technology Cooperation Program (TTCP) network in which RL will communicate via the satellite with the Canadians at the Communication Research Centre (CRC), NRaD will communicate with the Australians and CECOM with Great Britain [1].

The proposed experiment will consist of two phases. Under phase I experiment, the transmission of ATM cells will be performed at 45 Mbps between RL and CRC. Phase II experiment will be performed between RL and the Joint Interoperability Test Center (JITC), Ft. Huachuca, Arizona at 155 Mbps.

This report describes the program plan necessary to achieve these objectives. It documents the equipment and satellite requirements, data requirements, test procedures and all relevant mathematical calculations.

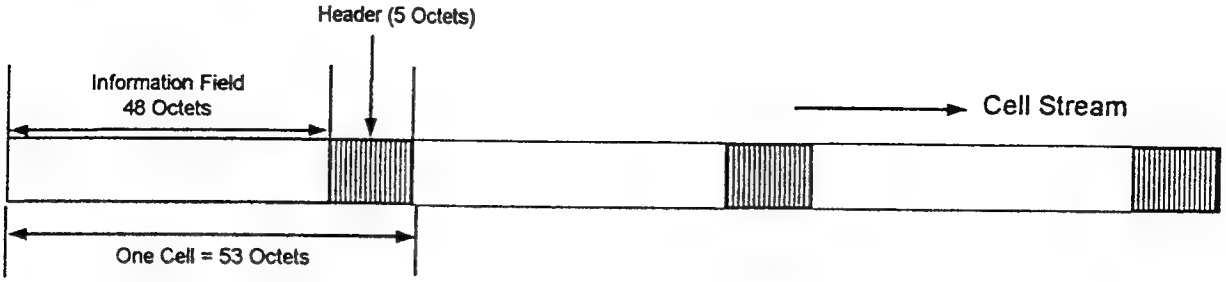
2 PRELIMINARIES

2.1 Asynchronous Transfer Mode (ATM)

ATM is a broad-bandwidth, low delay, packet-like switching and multiplexing protocol. It is a technique that uses constant-length packets as the basic units whereby the user information is carried in cells and is transferred asynchronously when it appears at the input of a communication system. It is essentially connection oriented whereby information is buffered as it arrives and inserted into an ATM cell when there is enough to fill the cell, and is then transported asynchronously across the network.

2.1.1 Basic Principles of ATM

Cell Structure:

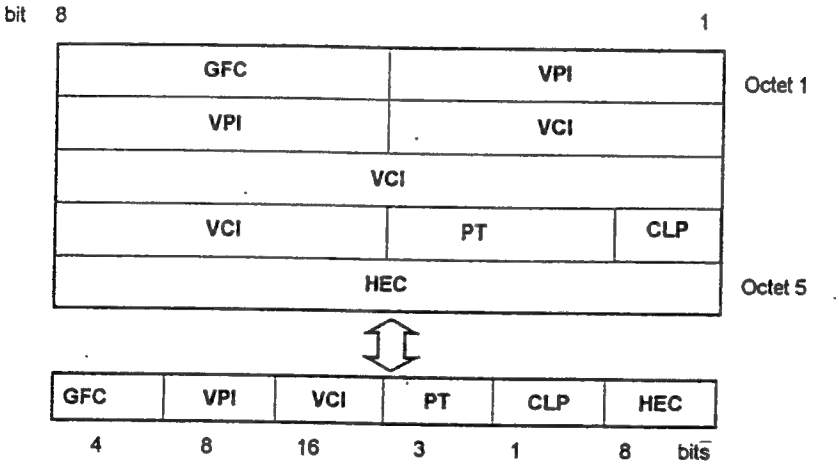


1 octet = 1 byte; 1 cell = 53 octets = 424 bits; Header = 5 octets = 40 bits.

Two header cell formats have been specified.

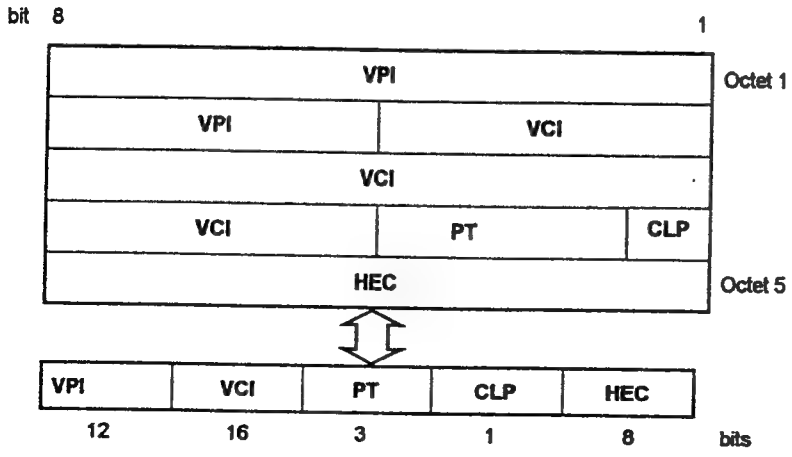
1. User-Network Interface (UNI)

   This header format is used for user-to-ATM connections.



2. Network-Node Interface (NNI)

   NNI is used for ATM-to-ATM connections.



where	GFC:	Generic Flow Control	VPI:	Virtual Path Identifier
	VCI:	Virtual Channel Identifier	PT:	Payload Type
	CLP:	Cell Loss Priority	HEC:	Header Error Control

In ATM, cells from more than one connection are multiplexed together, and at the receiving end the cell sequence is preserved. The information field (payload) is transported transparently (i.e. no error correction done) across the network by the ATM. Because it is a connection-oriented technique, a path must be established between the users before any information can be exchanged.

ATM Layers:

The ATM layer is concerned with transporting information across the network. It uses virtual connections for information transport and is divided into the virtual path and the virtual channel. The ATM layer performs functions such as generic flow control, cell header generation and extraction, cell virtual identifier translations and cell multiplexing and demultiplexing.

Due to the statistical behavior of the ATM network and the fact that a theoretical unlimited number of virtual connections can share the same link, it is required that an ATM switch synchronize to the incoming cell stream and examine the header field of each cell in order to identify the virtual connection and derive routing and control information for the switch network.

Another important aspect of the ATM protocol is the ATM Adaptation Layer (AAL). The ATM layer is independent of the services it carries and it relies on the AAL to adapt higher level data into a format that it can handle. AAL performs the necessary mapping between the ATM layer and the higher layers. Also, due to the payload independence, all the functions specific to the services are provided at the boundary of the ATM network and are performed by the AAL. Within the AAL, data flow can be corrupted by errors in transmission or it can suffer cell delay variation as a result of variable delay in buffers or through congestion in the network, thereby leading to cell loss or misdelivery of cells [2]. Since the services are dependent on the AAL type, it supports multiple protocols to fit the needs of the user.

2.1.2 Properties of ATM

Merits

1 ATM has the flexibility to match the rate of transmission to the rate of reception (i.e. variable bit rate). ATM can transport constant-bit-rate information at the same time as a variable-bit-rate bursty data. It does this by using a different ATM adaptation layer for each service being carried. Information can be transmitted through existing copper wire as well as through optical fibers.

2 The transmission of cells is matched to the generation of information such that it does not

transmit wasted cells during period of low multiplexing. This is known as statistical multiplexing.

3 ATM offers the ability of integrating various sources or media into one link. It can be applied to data, voice and video and may be used to integrate these applications.

Demerits

1 Cell delay variations (CDV)

This arises from the variable delays introduced into the network by the queues at the switches and multiplexers. This delay leads to a change in the expected gap between cells and is most acute in delay sensitive services such as speech transmission.

2. Cell Assembly Delay (CAD)

This arises because information from the source is buffered until there is sufficient information to fill a cell. The waiting time in the buffer depends on the arrival rate. This delay is also acute in delay sensitive services.

2.1.3 Theoretical Considerations and Practical Limitations

The theoretical analysis of the ATM systems is beyond the scope of this report. However, we will mention some important active research issues in ATM networks.

One of these issues is the cell loss performance of ATM networks. Cell loss probability has been estimated by model simulations using Monte Carlo and Importance Sampling simulation methods. The problem with these types of analysis is that the statistical behavior of the input and output queuing properties of the ATM switch has additional restrictions imposed on them. This is because the changes imposed by ATM queuing system introduces nonlinearity, large memory with random length, and complex traffic conditions. Thus a more suitable model (other than the conventional queuing theory) needs to be developed whereby the bias is applied to the conditional arrival probabilities.

Some of the practical aspects of ATM include the issue of generating and collecting ATM traffic using the adapter cards, the vulnerability of ATM signals to burst errors, and link BER characterization variation as a function of transmit power on the uplink.

2.2 Advanced Communication Technology Satellite (ACTS)

Spectrum congestion resulting from increased demand for telecommunications services have forced satellite system planners to consider operating at very high frequencies. ACTS is NASA's latest experimental communication satellite which was developed in order to research on how to accommodate this

projected increase in telecommunications demand in the future by effective utilization of the frequency spectrum. ACTS operates in the Ka-band (30/20 GHz) where atmospheric distortions, especially rain attenuation, are severe.

The on-orbit configuration of ACTS is shown in Figure 1 [3]. ACTS employs several elements not found in conventional bent-pipe transponders. Among many futures of ACTS, the following are more profound:

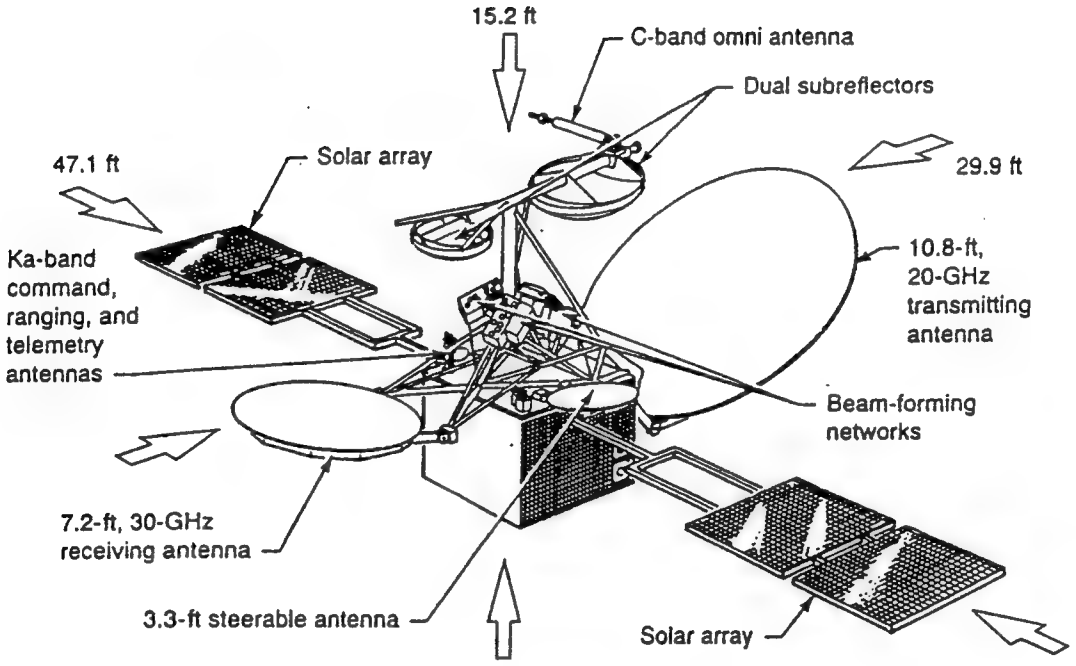


Figure 1: ACTS On-orbit Configuration

Multibeam Antenna (MBA)

ACTS employs multiple dynamically hopping spot beams. The MBA consists of two high-gain offset-feed Cassegrain antenna systems: one for uplink reception at 30 GHz band (2.2 m) with nominal gain of 60.2 dB, and the other for downlink transmission at 20 GHz band (3.3 m) with nominal gain of 56.7 dB.

Each antenna has five ports, two Baseband Forming Network (BFN) ports and three fixed-beam ports. Any part of the port can be connected to the Baseband Processor through the waveguide switches. These BFN ports are the East and West hopping beams receiving families. The operation of these two families are similar. Also connected to the West beam family is the steerable antenna which can be connected to any point in U.S.A. and Canada. It is possible to transmit and receive simultaneously through the steerable antenna, but it can only access one ground location at a time. If experiments are not co-located, using the steerable beam for uplink and downlink will not be possible. Also, Its main reflector is smaller and thus smaller gain is achieved when the steerable antenna is used.

On-board Switching and Processing System

There are two radiofrequency communication modes: Baseband Processor mode (BBP) and the Microwave Switch Matrix mode (MSM). In the BBP mode the received signal is demodulated, encoded or decoded if necessary, routed, and remodulated by the baseband processor and then retransmitted. Modulation for BBP must be serial MSK. BBP can receive modulated signal at two different transmission rates but output is limited to 100 mega symbols per second. It can accommodate 15 dB of uplink fade and 6 dB of downlink fade and still achieve bit error rate of 10^{-6} [3].

For the MSM mode of operation, the received signal are transmitted back without processing. The MSM is a programmable 4X4 switch that can connect a maximum of three transmitters to three receivers. It can operate in two modes: the static mode or the dynamic mode. The static mode is equivalent to the conventional "bent pipe" satellite mode of operation in which the switch configuration is fixed. In this mode, traffic can be FDMA or very high-burst-rate TDMA and bandwidth is limited to one GHz In the dynamic mode, traffic is switched around rapidly, and the modulation method is limited to digital modulation schemes, such as BPSK or QPSK. MSM mode can support 12 dB (8 dB) of uplink (downlink) fades and still achieve bit error rate of 10^{-6}.

3 TEST OBJECTIVE

Rome Laboratory is part of the Secure Survivable Communications Network (SSCN) program which has the objective (among others) of determining the feasibility, desirability and usefulness of ATM protocol as the communication backbone in a tactical environment. To accomplish this objective, RL will conduct series of tests over satellite links carrying ATM cells.

Satellites provide fast, low cost communication infrastructure deployable in a tactical environment where terrestrial network does not exist. ATM allows access to all types of media (voice, data, video). A combination of ATM and satellite provide an improved quality of service. This test will be an investigative test with the objective of verifying the above assertions. Each phase of this test will be further subdivided into two subtests.

3.1 Subtest I

Subtest I will be for the characterization of the satellite channel. Its purpose will be to establish the operating points of the channel and for link budget analysis. To specify the transmission quality of ATM cells over the ACTS at the rate of interest, the transmission performance obtained in conjunction with the Earth Station Terminals (EST) must be determined. In order to determine this, the carrier-to-noise ratio and the receiving system noise temperature in the presence and absence of rain is considered. This is because the most limiting factor in the Ka-band satellite transmission is rain fade.

Using the information on ACTS and the EST parameters, link budget calculations were made. These calculations include the uplink and downlink budget for RL node, the estimated uplink and downlink budget for CRC and the uplink and downlink budget for JITC.

3.2 Subtest II

Subtest II will be concerned with the application of ATM protocol over the satellite. To determine to what extent ATM cells can be transmitted over the ACTS, file transfer, video teleconferencing (VTC) and/or both will be used. Files (data files only) will be created and transmitted over the link. A copy of the transmitted files will be kept for comparison with the received files. Using the assets of the network group at RL and presumably at CRC and JITC, VTC will be established between any two nodes. The success or failure of the VTC will be noted and participating members will be asked to evaluate the quality of the services.

File transfer and video teleconferencing will be carried out simultaneously and if possible a voice transmission will be integrated. The objective is to determine the multimedia capabilities of ATM transmissions.

3.3 Data Acquisition and Analysis of Result

During link characterization, the cell loss rate, cell error rate and the link bit error rate will be determined or computed from measured data. The data collecting forms are given in Appendix B.

The transferred files are compared with the received files in order to determine the following:

a. Percent of transferred files without error and the percent of errored files.
b. For the errored files, the severity of the error.
c. Any correlation of error with the properties of the transmission data.

All relevant information must be recorded on data collection sheet.

For the VTC case, the quality of the video transfer must be determined. The participants are asked to subjectively evaluate the quality of service and record their observations.

An important objective in any experiment is to collect some information. A more important objective is to be able to interpret accurately the meaning of the collected data. In this experiment, we expect ATM cells to be lost, files to be corrupted and link BER to rise above expectations. This experiment must try to answer why these errors occur so that subsequent experiments will prevent or minimize the occurrence of these errors.

To achieve this objective, the results will be discussed with reference to, a) BER, b) causes of cell loss, c) discussion of limiting factors, and d) the correlation of data to error. The issue of generating enough data to fully load the channel should be discussed. And if the channel is not fully loaded, the exact rate at which experiment was performed must be specified.

4 SETUP AND CONFIGURATION

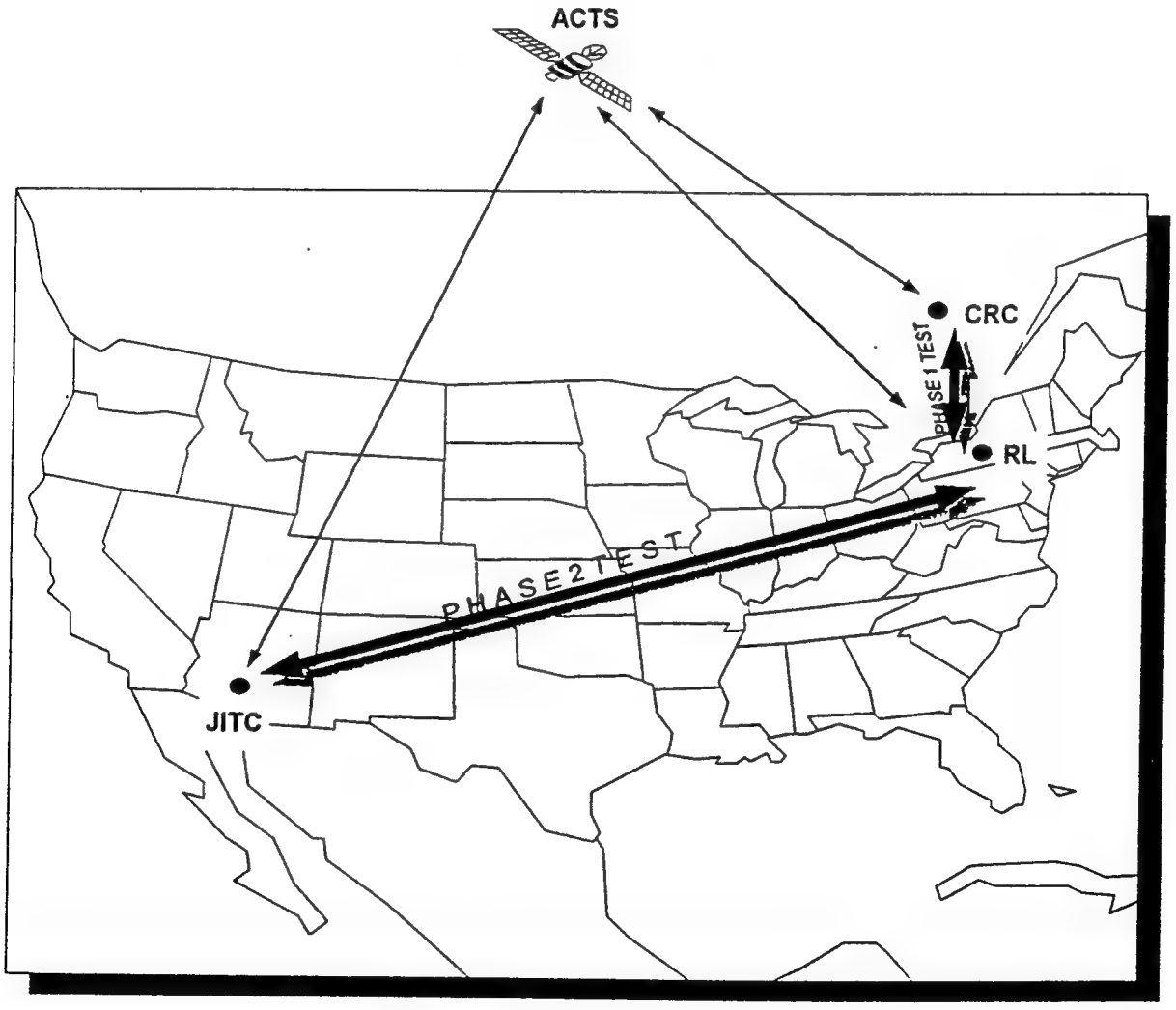
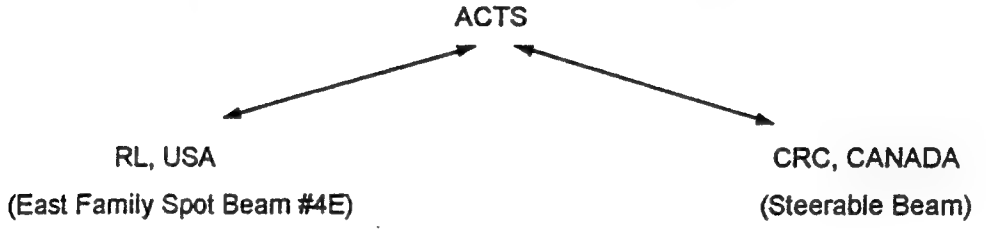


Figure 2: Experiment Setup (USA and Canada)

4.1 Experiment Description

Phase I experiment will link RL to CRC. The terminals at RL and CRC will transmit and receive ATM signal at DS-3 rates over the ACTS. CRC will provide the two terminals for this experiment. RL has a commitment from CRC to upgrade two of their old terminals (originally used with the Olympus satellite at DS-1 rate) to be able to operate at DS-3 rates. It is our understanding that these two terminals are upgradable to be compatible with the ACTS. For this experiment, we suggest the following scenario.

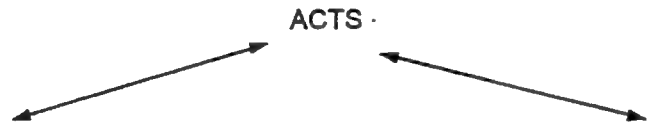


RL: Latitude: 43° 13' 13.176" N
Longitude: 75° 24' 34.003" W

Spot Beam 4E: Latitude: 42° 75' N
Longitude: 74° 86' W

CRC: Latitude: 46.05° N
Longitude: 77.20° W

Phase II experiments will link RL to JITC. The transmission rate for this experiment will be at OC-3 rates. The terminals (two required) will be provided by NASA. These are the HDR earth terminals offered by NASA to experimenters in conjunction with the ACTS. For this experiment, we suggest the following scenario.



RL: Latitude: 43° 13' 13.176" N
Longitude: 75° 24' 34.003" W

Spot Beam 4E: Latitude: 42° 75' N
Longitude: 74° 86' W

JITC: Latitude: 31° 134' 17" N
Longitude: 110° 20' 38" W

West Scan Beam centered at Phoenix, AZ.
Latitude: 30° 30' N
Longitude: 112° W

4.2 Phase I Experiment

The proposed end-to-end block diagram of phase I experiment is shown in Figure 3. As was indicated earlier, the satellite transmission equipments and the terminals will be provided by the Canadians. CRC is still working on the upgrade of these equipments and full information is not yet available. In the

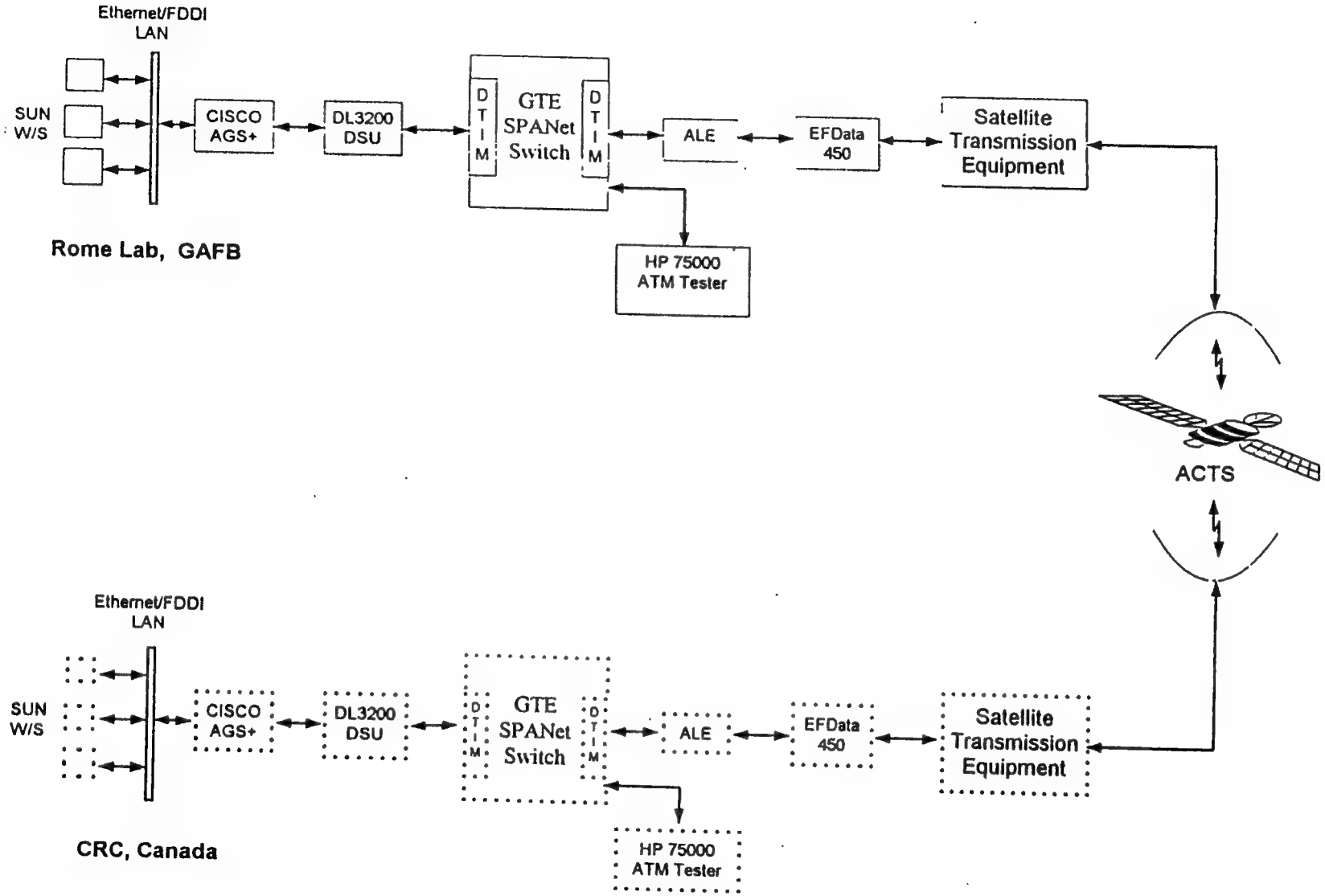


Figure 3: Phase I Experiment

absence of specific information on the terminals, we will assume the components shown in the Figure. The dotted boxes signify assumed components. Satellite data and application will be provided via the GTE SPANet ATM switch for the RL node. The SPANet switch at RL is already fully functional and equipped with DS-3 modules. No information is available on the ATM switch at CRC. Since full information is not available on the proposed Canadian terminals, we will assume a mirror image of RL equipments for CRC until such times when specific information will be made available.

4.2.1 ATM Switching Equipment

At RL the ATM generation and switching equipments shown in Figure 3 is the responsibility of the Network Group RL/C3BC. These components consist of the GTE SPANet switch, Device Service Units, Routers, LANs, and Workstations. These components are already installed and operable. These equipments are described below.

1. ATM Interface Hardware

The SUN workstations (W/S) at RL is equipped with ATM interface hardware to generate ATM cells. Each W/S is equipped with ATM adapter card.

2. GTE SPANet Switch

The GTE SPANet switch is the major ATM switching equipment at RL. The configuration of this switch is as follows [4]:

- DS-3 Trunk Interface Module (DTIM-3) with three 45 Mbps ports. Each port can access the Ethernet LAN and/or FDDI LAN.
- DS-1 Interface Module (DIM) with four 1.544 Mbps ports. Each port is access for T-1 carrier or TRI-TAC circuit switch through DTGI.
- Low-Rate Trunk Interface Module(LTIM) with variable rate port. Each port can access TRI-TAC radio and switches, satellite, T-1 carrier, and "Generic Low-Rate" devices.
- SONET Interface Module (SIM). The port is access to the Fiber Optic cable or the FDDI LAN at a transmission rate of 155 Mbps.
- Workstation Interface Module (WIM). This module provides access to workstations at a maximum speed of 100 Mbps.

3. Switch Accessories

The other accessories used in conjunction with the SPANet switch are the Device Service Units (DL3200), CISCO Router AGS+, and ATM cell converters.

4.2.2 Satellite Transmission and Communication Equipments

An ATM Link Enhancer (ALE) is recommended for this experiment. At high data rate, the ALE implements selective interleaving which significantly improves the performance of ATM traffic over satellite links. The ALE is capable of bit interleaving of the ATM header which improves the performance of the header error correction, thereby providing an acceptable "cell discard probability" [5].

The modem and other terminal equipments will be provided by CRC, Canada. Meanwhile, we will assume the equipment configuration as shown in Figure 3 for the Canadian terminals. The modem is assumed to be EFDATA 450 which has the capability of handling up to 45 Mbps. The satellite transmission equipment is assumed to be standard EHF satellite equipment.

4.2.3 Test Equipments

The HP 75000 Broadband analyzer is identified as the ideal test equipment and is recommended for this experiment. This piece of equipment is capable of generating ATM cells and monitoring the performance of the switch as well as the satellite channel. Different data on the ATM cells, the switch and the link can be collected with this equipment. This equipment is also vital for link characterization.

4.2.4. Test Procedure

Subtest I

1. Set up the equipment as shown in Figure 3.
2. Align the earth station antenna at both test nodes to the ACTS and set the frequencies and power levels as specified in the satellite access approval.
3. Initialize the test equipments HP 75000, and the ATM cell generating/switching equipments in accordance to the instructions contained in the operations manual or in accordance to the direction of the network branch.
4. Observe and record the status of all the equipment at time t=0 (initial point). Calibrate the equipment if necessary.
5. GTE SPANet switch setup instructions:
   - Please consult the network branch for setup and operation instructions.
6. Use the HP 75000 to generate and send ATM sample data for the subtest I. To do this, follow the procedure as specified below. Select "OK" or "Close Window" at the end of your selection.
   - Open the Test Session Manager and configure the system as follows:

     Set Adaptation: AAL-1

     All Others: Default values
   - Configure the test procedure

     Set the active port by activating the DS-3 interface.

- Open the Configure Interface window and set the following parameters:

  Select type of connection such that signal travel to and from the system under test (SUT) via the transmitter and receiver independently.

Framing Format:	C-bit
Interface:	UNI
Cell Scrambler/Discrambler:	OFF
Line Scrambler/Discrambler:	OFF
Header Correction:	ON
All Others:	Default Values

♦ Cell Generation

- Open the PDU sequence builder under the Builders window

  Under this window, select Create and give the PDU name as "Test_phase1"

  Select edit if necessary to edit your selection.
- Open the LIF sequence builder and create sequence from the ATM cells.
- Open the Foreground Cell Generation window.

  Open Set Content window (under Cell Access) and set the following:

VPI:	1
VCI:	1
PT:	0
Cell Loss Priority:	1
Defined Sequence:	Test_Phase1
Sequence Length	1500*
AAL Type:	1
All Others:	Default values

* This equipment can only generate a maximum of 1500 cells per second in the foreground condition and 16 cells per second in the background condition. This is not enough for 45 Mbps data. At this rate, the required number of cells is 106,132.08 cells per second. This is a problem that is still to be resolved. Filling the pipe with the HP 75000 is not possible!!!

- Open the ATM Cell Selector under Receiver window and set the following:

VPI:	1
VCI:	1
PT:	0
SN Correction:	OFF
AAL Type:	1

All Others: Default values

- Activate the monitor and capture cell stream in the Capture Memory window.
  - Open the Foreground Channel Generator and activate the transmit button.

  Select the display icon to view the captured PDUs. Store these PDUs for comparison with the received PDUs.

7. Data Collection
   - Use appropriate data sheets for data collection. Please indicate and record all observations. Use extra sheets if necessary.

Subtest II

To demonstrate the extent by which ATM can be transmitted over ACTS, high speed file transfer, VTC or both should be used to test the usefulness of ATM. For this demonstration set up the experiment as shown in Figure 4.

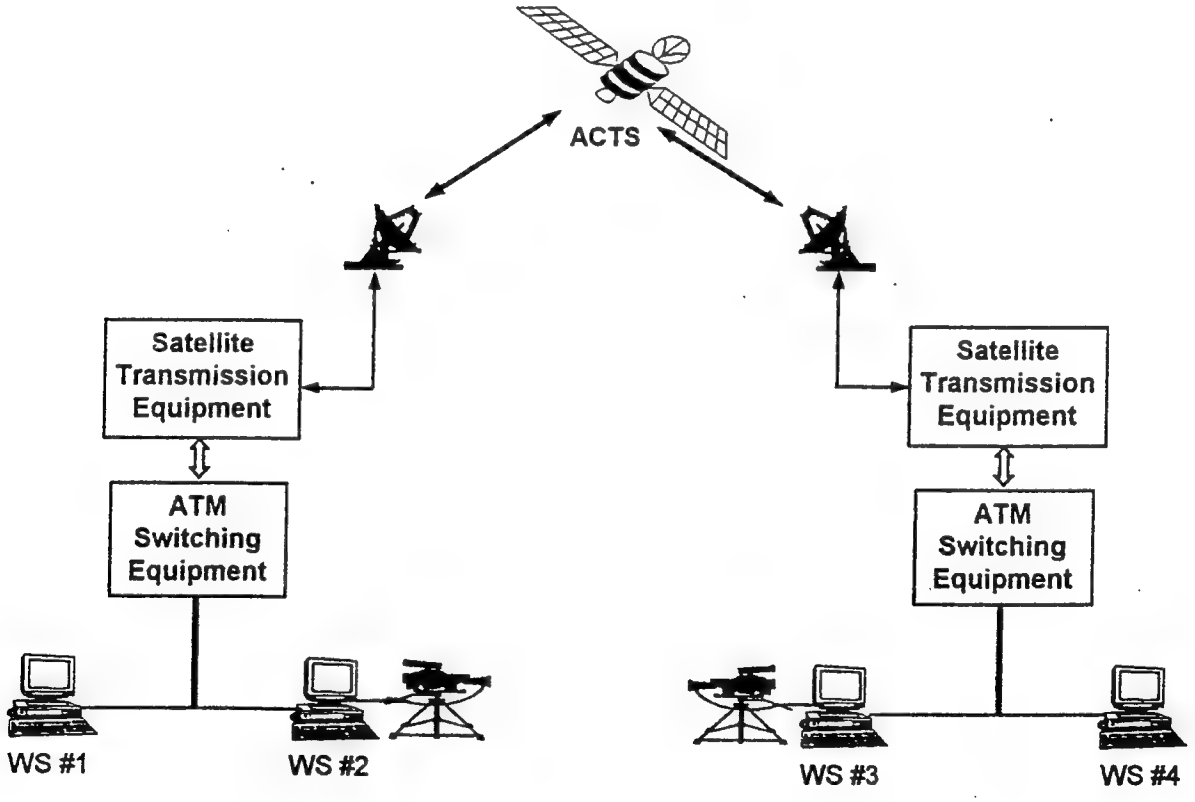


Figure 4: Application Experiment Setup

Workstations #2 and #3 should run a point to point VTC between RL and CRC. Setup and operation of the VTC is the responsibility of the Network Branch. Consult with the network group for direction on the VTC setup. Workstations #1 and #4 should be pre-loaded with a data intensive file (any file or combinations of files at least 45 magabyte). These files will be transferred between the two nodes.

4.2.5 Phase I Equipments

RL Node:

Description	On-Site Availability	Obtainable
Sun Sparc 10 Workstation	Yes	
ATM Computer Interface @ 45 Mbps	Yes	
GTE SPANet Switch & Accessories	Yes	
Video Teleconference Components - Camera with accessories - Tripod Stand - Speakers - Amplifier and Microphone	Yes	
ATM Link Enhancer	No	Yes (COMSAT)
Ka-band antenna with up/down converters, feed systems and accessories.	No	Yes (CRC)
EFData modem, SDM-450	No	Yes (COMSAT)
EFData Buffer EB-450	No	Yes (COMSAT)
HP 75000 Broadband Analyzer	No	Yes (Mitre)
Firebird Communication Analyzer MC6000	Yes	

CRC Node:

Description	On-Site Availability	Obtainable
Sun Sparc 10 Workstation	Yes ?	
ATM Computer Interface @ 45 Mbps	Yes ?	
GTE SPANet Switch & Accessories	Yes ?	
Video Teleconference Components - Camera with accessories - Tripod Stand - Speakers - Amplifier and Microphone	Yes ?	
ATM Link Enhancer	No ?	Yes (COMSAT)
Ka-band antenna with up/down converters and feed systems and accessories.	Yes	
EFData modem, SDM-450	No ?	
EFData Buffer EB-450	No ?	
HP 75000 Broadband Analyzer	?	Yes (Mitre) ?
Firebird Communication Analyzer MC6000	?	Yes (RL) ?

Note: ? is used to indicate unanswered questions regarding the equipments

4.3 Phase II Experiment

Figure 5 shows the setup of the test configuration for the phase II experiment. As we previously indicated, the experiment will be conducted between RL and JITC.

4.3.1 ATM Switching Equipments

As the block diagram indicates, both RL and JITC already have fully functional ATM switching equipments. Both sites have the GTE SPANet switch equipped with SONET Interface Module (SIM) operating at OC-3 rates. This switch has been described in section 4.2.1.

4.3.2 Satellite Transmission and Communication Equipments

The terminals for this phase of the experiment will be provided by NASA. This terminal is the 3.4m Ka-band transmit/receive HDR transportable terminal system assembled by BBN Systems and Technologies. These terminals will transmit and receive ATM signals at OC-3 rates via the ACTS between RL and JITC.

On the recent ACTS HDR experimenters meeting held in Cambridge Massachusetts at the BBN Systems & Technology complex, on July 6-7, 1994, the issues concerning the HDR experiment were outlined. These issues include the HDR Earth Station site requirement, HDR safety handling procedure, HDR terminal distribution allocation, ACTS time allotment, availability and acquisition of the HDR terminals. NASA, in conjunction with BBN Systems and Technologies, will provide all the necessary components and software to run the terminal. Also, the transportation of the terminals to and from the sites will be provided by NASA. However, the sites (RL & JITC) will respectively be responsible for the terminals site preparation, approval and the operation of the terminal after it is installed.

4.3.3 Testing Equipments

Since the connectivity of the terminal is not available at this time, we will assume that the HP 75000 and Firebird MC6000 will be compatible with the HDR terminal system.

4.3.4 Test Procedure

To stipulate the test procedure, we will assume that no extra-ordinary procedure is required for the HDR operation. If this is true, the test procedure will be identical to the procedure outlined in section 4.2.4 for the phase I experiment. The only difference is that this experiment will be performed at 155 Mbps as opposed to 45 Mbps. In other words, the terminals will be equipped with ATM/SONET user interfaces modules. Through these interfaces, this experiment will be performed via the ACTS.

In this phase of the experiment, we also encounter the problem of generating enough data for OC-3 transmission rate. At this rate, we require to generate 365,566.04 cells per second. The generation and testing of cells big enough to file the pipe is the bottleneck to this experiment. To the best of our knowledge, no equipment currently exist for the generation of such cells per second.

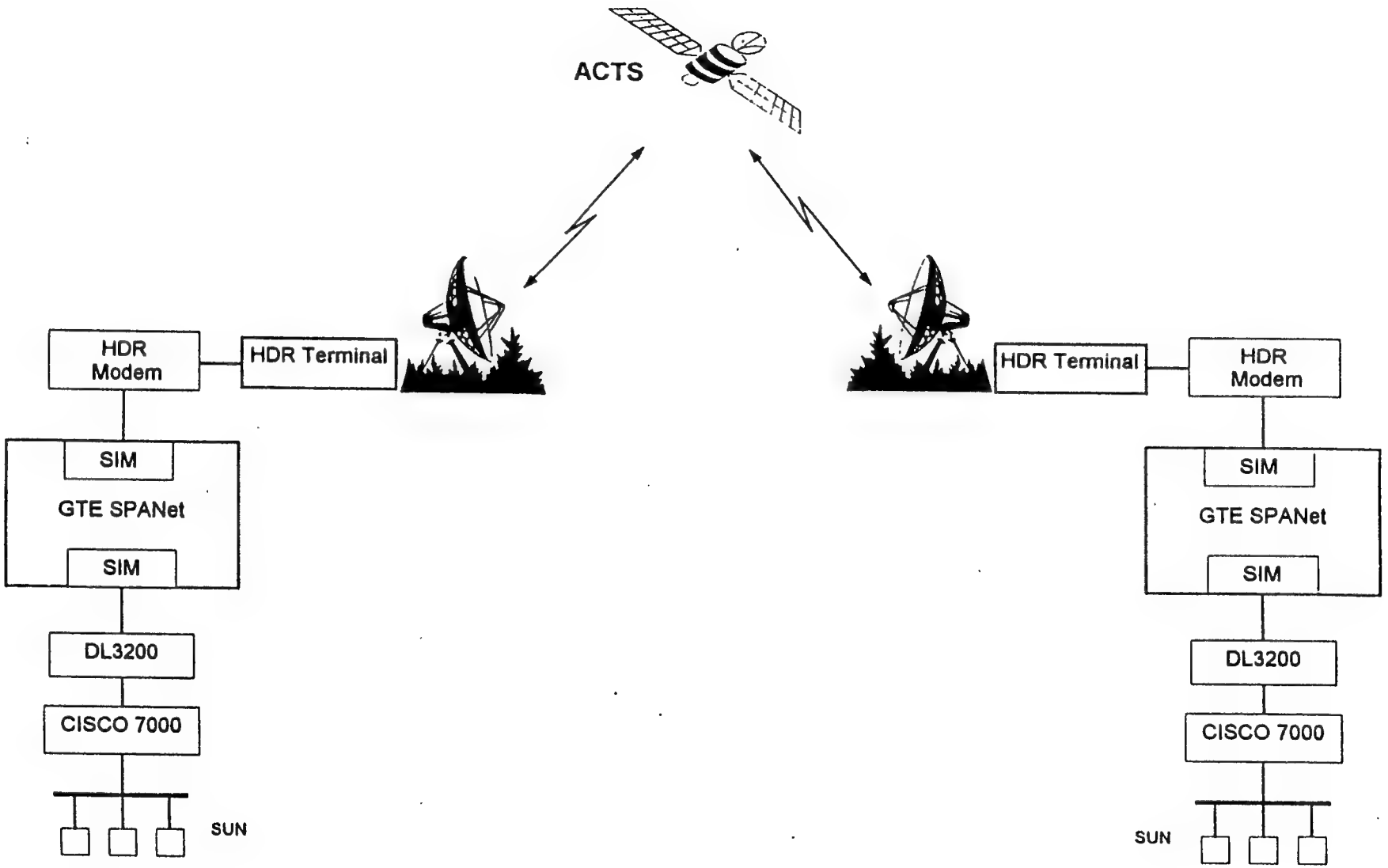


Figure 5: Phase II Experiment

4.3.5 Phase II Equipments

RL/JITC Nodes:

Description	On-Site Availability	Obtainable
Sun Sparc 10 Workstation	Yes	
ATM Computer Interface @ 155 Mbps	?	?
GTE SPANet Switch & Accessories	Yes	
Video Teleconference Components - Camera with accessories, Tripod Stand, Speakers - Amplifier and Microphone	Yes	
HDR Terminal with all the accessories	No	Yes (NASA)
ATM Link Enhancer	No	Yes (COMSAT)
HP 75000 Broadband Analyzer	No ?	Yes (Mitre) ?
Firebird Communication Analyzer MC6000	Yes	

5 SUMMARY

In this paper, we have described a program plan for transmitting high data rate ATM/SONET data over the ACTS. This plan is described based on the available information during the summer research program at the Rome Laboratory.

The execution of this plan is contingent upon several factors. Phase I experiment depends on the ability of CRC to upgrade their terminal to ACTS specification. Phase II depends on the availability of the HDR terminal from NASA.

6 REFERENCES

[1] G. A. Bivens, "Satellite Networking Research in Scalable Networking Technology, Rome Laboratory Technology Demonstration with NASA ACTS Satellite - (proposal)", May 1994.

[2] L. G. Cuthbert and J-C. Sapanel, *ATM The Broadband Telecommunications Solution,* England, Short Run Press Ltd., 1993.

[3] NASA Lewis Research Center, *Systems Handbook - Advanced Communications Technology Satellite,* Technical Report TM-101490

[4] N. Kowalchuk, "Secure Survivable Communications Network (SCCN), RL Project Report, 10/93

[5] COMSAT, *Demonstration of Asynchronous Transfer Mode (ATM) via Commercial Satellite,* Technical Report 10/25/93, COMSAT Technology Services.

[6] R. Manning, "A Unified Statistical Rain Attenuation Model for Communication Link Fade Predictions and Optimal Stochastic Fade Control Design Using Location Dependent Rain Statistics Data Base", International Journal of Satellite Communications, vol. 8, pp. 11-30, 1990

A STUDY OF THE APPLICABILITY OF FRACTALS AND KINETIC EQUATIONS TO ELECTROMIGRATION AND THERMALLY INDUCED HILLOCK AND VOID EVOLUTION

Joseph Chaiken
Associate Professor
Department of Chemistry

and

Martin Villarica
Graduate Research Assistant
Department of Chemistry

Syracuse University
Department of Chemistry
1-014 Center for Science and Technology
Syracuse, NY 13244-4100

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

August 1994

A STUDY OF THE APPLICABILITY OF FRACTALS AND KINETIC EQUATIONS TO ELECTROMIGRATION AND THERMALLY INDUCED HILLOCK AND VOID EVOLUTION

Martin Villarica, Graduate Research Assistant
Joseph Chaiken*, Associate Professor
Department of Chemistry
Syracuse University

Abstract

The growth of hillocks and voids in metal films was studied. The applicability of a model involving fractals and kinetic equations was examined on the basis of whether there is independent justification for using scaling arguments in the model and for whether there is reason to connect the evolution of hillocks with that of voids. Hillocks and voids were found to be self-similar across about three orders of magnitude of variation in spatial scale with the same fractal dimension. Voids and hillocks were found to have the same fractal dimension whether studied using atomic force microscopy (AFM) or scanning electron microscopy (SEM). The parameters obtained from these fractal analyses demonstrate quantitative internal consistency with an earlier time dependent study of thermal annealing effects on hillock distributions. Remarkably, area-perimeter data obtained from either a long-time study of a single void or a spatial everage of a large number of different voids both yield quantitatively identical results.

A STUDY OF THE APPLICABILITY OF FRACTALS AND KINETIC EQUATIONS TO ELECTROMIGRATION AND THERMALLY INDUCED HILLOCK AND VOID EVOLUTION

Martin Villarica and Joseph Chaiken*

Introduction

We recently reported[1] the very encouraging results of our first attempt at applying a kinetic model[2] to the time evolution of the growth of hillocks in metal films. These results raise various questions requiring satisfactory exploration before we can confidently form conclusions regarding any possible relationship between the assumptions and parameters of the model and actual physical structures and interactions. Fundamental to the model are assumptions and parameters derived from the scaling behavior of the microscopic parameters describing the entities responsible for hillock and void growth. Additional relationships between the spatial, temporal and mass scaling of the equations describing the interactions between those entities also require exploration. In what follows we describe our attempt to obtain independent evidence justifying the scaling assumptions of the model and their application to electromigration.

First, we will briefly review the model and the results of our application of the model to data obtained independently by Vook and coworkers[3,4]. Naturally, the questions we hoped to address this summer are articulated in the succeeding section. We then present the results of atomic force microscopy(AFM) and scanning electron microscopy(SEM) experiments on aluminum films which provide some answers or at least insights into these questions. It should be clearly noted from the outset that what we seek is a general model for rearrangement of thin film matter when subjected to thermal and/or electromigration stress and so much work will be needed to establish the reality of the physical picture suggested by any model. We believe the potential empirical use of the model for reliability and failure prediction will be clear in any case. Whereas the log normal distribution has been applied to electromigration in various ways, our model employs the Poisson distribution which usually fits experimental data equally well or better. Our model assigns physically meaningful quantities to the microscopic parameters of the Poisson distribution which, at worst, could be the basis for a semi-empirical parameterization. The log normal distribution cannot offer the same opportunity for connection with microscopic causes and effects.

The Model

The model has been described in detail in the literature[2,5] and is based on the Smoluchowski equations **(1)** with Jullien's asymptotic solution[6] **(2)** giving the time dependent concentration, n_k, of a cluster, hillock or some other association or entity having k monomers units, i.e. a k-mer. The model assumes that k-mers are only produced by coalescence of smaller clusters and are only destroyed when they coalesce with another cluster to form a still larger cluster. If there was a population of monomers at t=0, say metal atoms impinging on a substrate from the gas phase, the formation of a film would commence by the formation of associations between the atoms. These "associations" might be thought of as islands, hillocks or some other name depending on their specific properties.

$$\frac{dn_k}{dt} = \frac{1}{2}\sum_{j=1}^{k-1} K_{j,k-j} n_j n_{k-j} \; - \; n_k \sum_{j=1}^{\infty} K_{j,k} n_j \qquad (1)$$

$$n_k \rightarrow C\, k^a\, e^{-bk} \qquad (2)$$

Although this model seems to contain much of the physical nature of the hillock/void formation and growth process, realistically, we know there are at least a few potential processes which could lead to fragmentation of some islands or clusters. Although both unimolecular[7] and bimolecular processes[8] can be imagined, we will group them all together under the heading of "evaporation". There is no explicit term for evaporation at this point except to note that the rate constants used in **(1)** implicitly relate the fact that not all encounters between i-mers and j-mers result in coalescence. The greater the probability of evaporation occurring on some time scale related to the averaged time before the next collision occurs, the smaller the crossections or rate constants used in **(1)**. Since in this model the presence of a buffer system, e.g. buffer gas in gas phase systems or codeposition gas in matrix isolation, does not appear explicitly, these rate constants contain this information implicitly when the equations are fit to real data. The asymptotic solution, **(2)**, can be interpreted in terms of the scaling of the bimolecular rate constants, $K_{i,j}$, relating the interaction of individual agglomerating entities, e.g. i-mers and j-mers, with increasing numbers of monomers in the clusters. Under certain situations[9], it is thought that the rate constants could scale smoothly with increasing the number of monomer units, and if **(3)** is satisfied they are termed "homogeneous".

$$K_{\lambda i,\lambda j} = \lambda^{2\omega} K_{i,j} \qquad (3)$$

The most empirical mathematical implications of homogeneity and scaling are best summarized by noting that they allow the rate constants, the $K_{i,j}$, to be written explicitly in terms of i and j. This allows analysis of the equations and the possibility of discovering rigorous closed form mathematical connections between microscopic parameters defining the $K_{i,j}$ and actual measurements of hillock or cluster size distributions.

Although we shall choose not to do so in this work, it is possible[10] to factor the $K_{i,j}$ into a product of two functions. The first involves only those factors which calculate the probability of a hard sphere(or other generic shape) encounter. The second is a probability factor relating the likelihood of an encounter ending in a certain action, e.g. coalescence. The first factor is essentially the collision frequency, i.e. related to overall mass transfer conditions, while the second contains information related to appropriately averaged trajectories and so contains the chemistry of the each reaction system in the agglomeration process. For very small clusters and entities, the observation of quantum effects is expected in the form of discontinuities in the variation of the $K_{i,j}$ with size.

In the absence of the above factoring of the $K_{i,j}$, the model's results, e.g. a distribution of cluster sizes produced in the long time limit, show what would happen if the distribution function was determined purely by kinetics. These kinetics would be determined entirely by mass transfer as if there were no preferred pathways for the formation or destruction of any size satisfying the homogeneity condition. The Stockmayer formula[11] from polymer chemistry relating the number of independent kinetic pathways producing a given size cluster to the number of monomers in that product cluster plays a prominent role in the rigorous proof of the validity of **(2)**. We assume that neither the total number of monomers, whether free or in clusters, nor the temperature, varies with time. The model is a so-called "mean value" model in that there is assumed to be no spatial correlation between the locations of the various reactants and products.

These scaling laws[9] are consistent with specific limits of attainable reaction conditions. They are based on rigorous calculation of things like how the mass of a cluster increases with increasing number of monomer units while the velocity of the clusters decreases. The scaling of mass transfer and energy transfer, in addition to reaction rates, depends on such quantities. The scaling of surface area of a cluster with increasing number of monomers is another such quantity. The distribution of the mass of the clusters with increasing monomer size is yet

another. Assumptions of Jullien[6,9] and others applying the Smoluchowski equation embrace specific ways of mathematically representing these quantities so as to make the equations describe systems which are "self-similar" across spatial and temporal scales. We shall define the term "self-similar" in this connection more precisely a little later.

We assert that the more often a given cluster encounters other clusters, the more rapidly it coalesces with other clusters. As a cluster gets larger, it becomes more massive, therefore moving more slowly, thereby exploring space more slowly and encountering fewer potential coalescence partners per unit time. The result is a slower growth rate. Balanced against this effect is the fact that as monomers or other size clusters are added to a cluster, it must get larger so its cross section increases. Because of this, more space is explored per unit time and so a faster growth rate obtains. The overall effect depends on whether a cluster becomes spatially larger faster than they become slower due to becoming more massive. The validity or importance of this physical picture in determining size distributions and other observable characteristics of coalescence growth systems is one of the fundamental questions regarding the entities responsible for hillock and void growth mentioned in the Introduction.

It is possible to show[6] that these effects, as well as the scaling of concentrations with varying spatial scales, are connected by **(4)**. This equation can be shown valid in a variety of commonly attainable sets of reaction conditions.

$$2\omega = \alpha + (d - d_w)/D \qquad (4)$$

(4) is obtained by considering the relationship between the spatial, temporal and mass scaling of the Smoluchowski equations. Mathematically, one considers a microscopic collision frequency or rate constant and the effect of increasing(or decreasing) each length which enters into the microscopic calculation of the rate constant or in expressing the concentrations. Calculating the reciprocal temporal scale change needed to make the terms of the Smoluchowski equations invariant to the net effect of the two scale changes, one encounters **(4)** as a necessary condition. We do not know whether other forms of the kernels exist which also produce self-similar distributions. What is important for this effort is that Jullien's solution does[6].

The first parameter, α, is obtained due to the scaling of velocity with increasing cluster mass. It will usually be negative. For an isothermal system at thermodynamic equilibrium, heavy objects move more slowly than light ones.

$$V_{rms} = \sqrt{\frac{3kT}{m}} \qquad (5)$$

(5) is simply the root mean square speed of an ideal gas molecule of mass m at temperature T. V_{rms} obviously scales as $m^{-1/2}$ so when the mass transfer is limited by the molecular speeds and not the concentration gradient, i.e. the ballistic regime and not the diffusive, $\alpha = -0.5$. The parameter d_w, the "fractal dimensionality of the cluster trajectories", results from the scaling of the mean square displacement between collisions with increasing cluster mass. When motion is diffusive, $d_w=2$, when ballistic, $d_w=1$. The dimensionality of the space in which the coalescence occurs is d. For clusters formed in bulk gases, d=3. For islands formed on single crystal surfaces by interaction of adsorbates diffusing along the surface, d=2. The fractal dimensionality of the clusters, defined in terms of a mass fractal, is D. These parameters define, and provide quantitation of, energy and mass transfer properties of the reacting system and chemical properties of the clusters.

When the shape of an experimentally produced cluster size distribution can be quantitatively calculated, based on scaling properties and the assumption that the kernels are homogeneous, we accomplish an enormous decrease in the number of variables needed to model the system. There are physical implications as well. For cluster or hillock sizes corresponding to homogeneous kernels, there are no special structures or interactions that occur as the clusters become larger. Alternatively, it may also be true that a given sized cluster occurs in various equivalent structures. This is consistent with the idea that the Smoluchowski equations produce self-similar cluster size distributions. It has been stated[12] that surfaces produced from distributions of self-similar hillocks and voids are self-similar, i.e. fractal.

Application to Vook's Data

Vook and coworkers allowed us access to several hillock size distributions that they produced using evaporated aluminum and aluminum-copper films on silicon substrates. Obtained using SEM, these distributions corresponded to a nominal magnification of 10,000x. The films systematically surveyed a range of deposition conditions, i.e. substrate temperature and deposition rate. The data were well fit using any of a few different distributions. Overall, the Poisson distribution usually fit best for pure aluminum films while alloys were often fit slightly better by a log normal distribution. Based on the quality of the fits it could certainly be said that homogeneous kernels could correspond to the physical situation represented by these

films. Vook approximated the hillocks as non-overlapped hemispheres; his raw data consisted of lists of diameters.

Vook's data suggested that SEM at 10,000x magnification is not adequate to discern the smallest and most mobile participants in the coalescence process. The distributions therefore poorly reflected the value of a from **(4)** that could be obtained from the fits to **(2)**. The value of b from **(4)** should be reasonably well determined. In fact, the b parameter is valuable because a plot of ln(b) vs ln(time) for the time evolution of a given system should be linear with slope $1/(2\omega-1)$. Since $a=-2\omega$, having such data allows a test of internal consistency for the application of the model to a given system. In the case of Vook's data, the value of a obtained from the plot of ln(b) vs ln(time) was -0.25 and comparable to the average value obtained from all the fits to the Poisson distribution directly, -0.21. On the basis of this and other results the application of the model was very encouraging.

The Summer of '94

There were many potential goals which we could have pursued this summer that would have been intelligent extensions of our work with Vook's data. We decided to proceed along a few different tracks in order to hedge against the possibility that various potential difficulties would defeat one or more of our efforts.

We hoped to be able to duplicate Vook's time dependent annealing experiment using a pure metal film, preferably Al, with greater resolution. We were able to obtain a time dependent data set on a pure aluminum film. The magnification was only 5,000x and the number of hillocks found was in fact less than the number obtained by Vook. Our statistics cannot be as good as Vook's data and we will still have a poorly determined a parameter from the direct fit of the Poisson distribution function to the experimental data, but we will be able to obtain a plot of ln(b) vs ln(time). When completed, we will have an a value which can be compared with that obtained by us using Vook's data. There is a real possibility that we can combine this data collected by Dr. Walsh with some data collected by Jennifer Synowczynski which will improve our statistics. This data is still being analyzed and the results cannot be discussed further. There is no question that we will ultimately want to apply this model to films of various metals on a variety of different substrates.

It would clearly be useful to obtain data at much higher resolution so that an accurate and precise a value can be determined directly from the fit to the Poisson distribution. We hoped to

be able to obtain independent evidence that scaling arguments and scaling based models have relevance to hillock and void evolution in thermal and electromigration stressed films. It would also be important if we could establish any quantitative and preferably analytic, direct correlation between hillock evolution, to which our kinetic model pertains, and void evolution which is clearly of central importance in electromigration induced device failure.

We decided to attempt to establish the proposition that voids and hillocks in metal films are self-similar. Establishing this as fact would establish the relevance of scaling in the analysis of hillock and void distributions. Using atomic force microscopy(AFM), it would be possible to establish self-similarity and to examine the hillock distributions at much higher resolution than is possible using SEM on these types of samples. It should be pointed out that SEM studies cannot be performed using conductive paints or other approaches to keep charging effects from limiting the resolution since any approach which contaminates the surface is not compatible with time dependent studies.

Self-Similarity

The question of exactly what physical meaning, if any, the term "self-similarity" denotes is deserving of some attention at this point. Within the context of having limited space, we begin with an example of a generic mathematical definition paraphrased from Korvin[12]. We consider a physical structure, perhaps a hillock in a metal film, the Nile river, or the Andromeda galaxy. It is possible, for whatever reason, to consider various ways to characterize the object's shape. In the effort to do so, we might consider a single object or set of objects, that for independent reasons we suppose might have something to do with each other, and the set's distinguishing characteristics as compared with other sets of objects.

Attempting to be exhaustive in our considerations, we ask whether there exists a relationship between the area of an object, or a 2-dimensional projection of the object, and the perimeter of the object. The comparison is based on the idea that a perimeter increases linearly as larger objects in a self-similar set are considered, whereas the area increases quadratically as larger members of the set are considered. "Larger" might be based on any convenient measure such as the largest distance across the interior of the object, the perimeter, or the area. Since we have been careful to make no assumptions regarding the actual shape of any objects we may consider, the ratio of these two quantities might be sensitive to the shape of the objects.

We can express the idea of considering the value of one measure of size and shape, say w, in terms of the simultaneous value of some other measure of size and shape, say L in terms of the equation below.

$$w=W(L)$$

Now if this relationship holds for all the members of the set, then it may be possible to write an equation which connects the members of the set in terms of the scaling of perimeter or area, i.e. L. Furthermore, if it is possible to write **(6)** below in which f is an unknown function which depends on the shape of the objects. This form amounts to a very significant assumption because if a set of objects can be found which satisfy **(6)**, then **(7)** can also be shown to be true and the objects form a self-similar set.

$$w_2/w_1 = W(L_2)/W(L_1) = f(w_2/w_1) \qquad (6)$$

$$f(ab)=f(a)f(b) \qquad (7)$$

It is then possible to apply Cauchy's treatment of functional forms to write that the function f connecting values of the two measures must be

$$w = w_0 L^p \qquad (8)$$

of the form above. At first, this may seem like a real leap forward in that we have a functional form which at the very least allows us to classify sets of objects. The use of group theory to classify various matrix elements by their behavior with respect to their behavior under various point and space group transformations seems analogous. Empirically, self-similarity and point group symmetry are analogous in that they allow a reduction in the number of variables needed to specify the state of a given system but in the case of self-similarity we presently have very little with which to satisfy our intuition as to why the classification should be physically meaningful.

First, there is nothing in the definition of self-similarity, i.e. satisfying the equations above, which pretends to allow a general first principles calculation of the prefactor f or p. Each case must be considered as an isolated situation. The significance and physical interpretation of the exponent p, known in the vernacular as a variety of different kinds of

fractal dimension depending on the specific case under consideration, beyond the idea that area or perimeter increases as an object gets larger, is usually difficult to clearly discern. The fact that we have chosen to express these measures of size and shape in terms of each other and examine their scaling behavior is not guaranteed to have any particular significance. To state one overall benefit of using self-similarity we directly quote H. E. Stanley[13].

> "If you are an experimentalist, you try to measure the fractal dimension of things in nature. If you are a theorist, you try to calculate the fractal dimension of models chosen to describe experimental situations; if there is no agreement then you try another model."

In addition to having discovered an empirical measure that a correct theory of whatever processes being considered must correctly reproduce, if one has a theory which predicts self-similar objects, then discovering that they exist in reality clearly supports the model. If a model *assumes* self-similarity and makes predictions which turn out to be true, then this is good but we are still not guaranteed that the model is unique. There could be another model not based on self-similarity which also accounts for the empirical data. However, discovering that the objects are actually self-similar clearly supports the theory which makes that assumption. Our hillock model expresses the mass distribution in terms of a mass fractal, i.e. self-similarity is assumed. Another important benefit is that self-similarity allows construction of a models having relatively few adjustable parameters. Therefore we attempted to find our whether hillocks and voids under thermal annealing or electromigration conditions are self-similar.

Again, following Korvin[14] and Stanley: "The numerical determination of fractal dimension of an object embedded into the d-dimensional Euclidean space is generally based on the following simple rule: Suppose we plot on log-log paper some quantity that can be interpreted as 'number of objects' or 'mass' against a characteristic length scale L. If there is an asymptotic ($L \to \infty$) region in which this plot becomes straight, then the slope is termed the fractal dimension d_q characterizing the distribution of the quantity q. Alternately, we can keep the size of the object, L, fixed and determined q for different resolutions, l. If we plot q against l, and the plot is a straight line in a range ($l < l_0 << L$), the fractal dimension will be the negative slope of this line. For example, if q is the area of a square of side L, it scales with L as L^2, if is the number of squares of side l<<L contained in the large square, it scales as:"

$$q=(L/l)^2 \ \alpha \ l^{-2}$$

Of course, this all follows from **(6-8)** above.

When examining surfaces, there is considerable latitude in defining hillocks and voids. The AFM data consists of a set of points, x, y, z defining the surface contour of the sample. We choose to view the surface as a collection of hillocks and voids because we believe the evolution of hillocks parallels the evolution of voids. The fractal dimension of a set of hillocks is conveniently calculated using a so-called "area-perimeter plot". Using **(6-8)** above it is easy to show that a plot of the logarithm of the area of a set of hillocks as a function of the logarithm of the perimeter is linear with slope D when the hillocks are a self-similar set. D is the fractal dimension of the projection of the hillock on 2 dimensions. The fractal dimension of the surface composed of the hillocks or voids is commonly taken as D+1 although, according to Korvin[15], the rigorous proof of this has not been given. Given AFM data, one need only choose some plane of reference in order to designate all the surface structures contained in the image as either hillocks or voids.

The choice of reference plane requires some discussion of thresholding. The basic fact of scanning probe microscopy is that some type of small probe is brought into close proximity to an object under study. As the probe, which is a very fine tip of one sort or another, approaches the the object to within a distance on the order of molecular dimensions, a variety of forces between the surface and the tip become measurable. By sensing one of the processes driven by these forces, e.g. electrical conduction, van der Waals attraction or repulsion, it is possible to sense how close the tip is to the object. This allows the mapping of the surface contours of the object as the tip is rastered across the region of interest at constant height. Balancing these forces allows assignment of a z value to each x,y point on the surface. "Thresholding" refers to construction of iso-z contours in the x,y, plane. The contiguous regions of x and y that have larger or lower z than the threshold are designated as hills or voids respectively.

A useful analogy is that of visualizing the flooding with water of a hilly region. The depth of the water determines the number of visible hills and lakes. A high threshold corresponds to deep water. Figures 1a-b are AFM images showing that metal films are very rough at all scales. For the 1μm square region (resolution of 500 x 500 lines) scans, the lateral resolution could approach 20Å but we suspect the tips are much wider. The z resolution is on the order of 1Å. The smallest features are on the order of 500 Å in the xy plane and most of the voids(lakes)

have the same shape near their bottom. Since it is arbitrary how we define the zero point level for the z direction, we can create hills or voids arbitrarily by our choice of z=0 elevation. Some hills are not of higher z than the water level. Analyzed from this point of view, the question of thresholding becomes a percolation problem involving the formation of a single contiguous flooded region or land region. The number of voids and hills is a strong function of threshold although it is thought that the fractal dimension calculated for a particular choice is much less sensitive if chosen far from the percolation threshold.

Figure 2 shows a graph of the number of hills and lakes found as a function of z threshold for a particular AFM image which is typical of those studied. The film was 7000Å of evaporated aluminum on silicon. Both substrate and film were passivated so we suspect that our images are obtained through a conformal oxide layer on the order of 20-30Å thick. The thresholding is normalized such that the limits of the abscissa are 0-100% of the maximum z found for that image. Figures 3a-c show graphs of the surface fractal dimension calculated for a range of thresholds for a few different films. Figure 4 shows a series of typical area-perimeter plots based on either lakes or hills obtained from another image. These results reveal much about the structure of these films.

Figure 2 suggests that a cut into a film, along a plane perpendicular to its surface would reveal a profile like that shown in Figure 5a in which hillocks are taken as spherical. The point here is whether the bases of different sets of adjacent hillocks make contact at different z values or whether a picture more like Figure 5b is appropriate in which there are very flat regions separating isolated hemispherical hillocks. The situation depicted in Figure 5b would go from zero hills to the maximum found within the smallest z increment instrumentally possible and then decline in numbers monotonically for increasing z thereafter. When there are hillocks on top of hillocks, increasing certain z thresholds will cover more hills than they will separate. This results in the number of hills or voids found not immediately being a maximum value at the lowest z value for which hills are found.

In the midwestern United States, a favorite country witicism is: "For every hill there's a hole." The land is so flat that the statement is very often true in actual fact. For every film studied there was always a choice of z threshold possible for which the number of hills was equal to the number of voids. This was always very close to the percolation threshold. It is clear that there is not perfect symmetry between the number of hills or voids as a function of threshold. It is also clear that the fractal dimension is a function of z threshold and that it

sometimes seems as if there is a systematic variation occurring. This deserves more study as does a more quantitative study of the variation of the number of hills and voids with choice of z threshold.

The variability may be unfortunate if we eventually want to establish a reproducible protocol for thresholding that will allow calculation of fractal dimension that can be compared from sample to sample, as a function of electromigration or thermal annealing or perhaps after various other types of film processing. While unfortunate, it still must be admitted that the entire range of surface fractal dimensions found for the AFM images of aluminum was roughly 2.57±.11 regardless of whether voids or hillocks were used in the area-perimeter plots. If a systematic part of the variability exists and can be quantitatively analyzed then we suspect that the uncertainty on the fractal dimension measurement could be reduced by an order of magnitude or more. This would greatly increase the potential utility the observable fractal dimension for various applications. Although we were only able to obtain a few usable images, it should be noted that for laser deposited platinum films on quartz, the fractal dimension found from area-perimeter plots was 2.55 ± .09.

Graphs illustrating the other type of analysis mentioned by Stanley involving changing the size of the region observed and counting the number of objects are shown in Figure 6. The slopes of these graphs for the aluminum films give a surface fractal dimension of ≈2.55 while for the platinum film we obtain 2.9. Why the aluminum films seem to give better agreement with the value obtained from area-perimeter plots compared with the agreement displayed using platinum film we cannot say. The fact that the fractal dimension obtained using hills is not distinguishable from the value obtained from voids supports our hope that we may be able to connect the behavior of hills predicted using our model with that of the voids.

These fractal dimension measurements can be used as a check of the quantitative consistency of our coalescence growth model. To utilize **(4)**, we assign $\alpha = -.5$ and $d_w = 2$. This follows from assuming that hillock growth occurs by diffusive motion of very small clusters and that the velocity of the mass transfer scales as the inverse square root of the mass of the cluster, consistent with the usual simple kinetic theory. The fractal dimension of the surface of the of the films, which may be where most of the material involved in coalescence growth can be found just before coalescence events occur, is $d \approx 2.55$. This number therefore corresponds to the dimension of the space in which the coalescence growth is occurring. Finally, the dimensionality of the hillocks themselves is also obtained by adding unity to the slope of the

area-perimeter plots giving D = 2.57. Using **(4)** we predict $\omega \approx -0.29$ which can be compared quite favorably with $\omega \approx -0.24$ from Vook's ln(b) vs ln(t) plots.

We obtained perimeter-area plots using SEM data for comparison with the same kinds of plots based on AFM data. Vook's data was not available for this purpose since he only reported diameters as his raw data. For this purpose, Professor Kime[16] allowed us to analyzed the data she had obtained last summer. She obtained area-perimeter data on over 1000 voids of similar passivated aluminum films. The plot obtained from this data is in Figure 7 and, remarkably, over 900 voids form a reasonable line giving a surface fractal dimension of 2.71 in good agreement with the values obtained from AFM data. Many of the voids which deviated significantly from the line shown were in fact on the lower limit of the resolution of the instrument. This is a common pitfall causing the measured areas to depend of the instrumental resolution and that data would have had to be discarded on that basis regardless of any other considerations.

Finally, we utilized the remarkable data of Thomas and Calabrese[17] involving the evolution of a single void under electromigration conditions. A single void was observed under nominally constant conditions for a period of hours with images being obtained every few minuted or so. This data was converted to a standard area-perimeter data set as shown in Figure 8except that all the areas and perimeters correspond to the *exact same void.* The plot yields a reasonably straight line giving a surface fractal dimension of 2.67. It appears that the long time average behavior of a single microscopic void can be replaced by a spatial average of a large number of voids at a single time. We believe this to be a singular result in science. One of the basic precepts of statistical mechanics is explicitly confirmed by direct observation.

This final result is profound. It again shows that the fractal dimension of voids and hills are nearly equal, that AFM and SEM data are comparable, that the fractal dimension of the voids, and therefore probably hillocks, *remains constant throughout their evolution.* Despite the fact that the void changes its exact shape repeatedly in their data, the fractal dimension could be used in **(4)** to follow the growth of a distribution of such hillocks(voids).

Summary and Conclusions

At the precision available to us, voids and hillocks are self similar across spatial scales spanning greater than three orders of magnitude and they appear to have the same fractal dimensionality. Fractal exponents obtained from either AFM or SEM images can be inserted into

the equation governing the scaling parameter ω from our kinetic model of hillock evolution. Direct observation shows that the fractal dimensionality of an electromigration driven void does not vary with time during its growth. The fractal exponent for passivated aluminum metallization patterns is $2.57 \pm .22(2\sigma)$.

Acknowledgements

I would like to acknowledge the beneficial interaction and assistance provided by Profs. Yolanda Kime and Herbert F. Helbig, Drs. Joe Beasock, Lois Walsh, Tom Renz, and G. Ramseyer, and the summer students Jennifer Synowczynski and Nathan Terry.

References

1 J. Goodisman, J. Chaiken, Thin Solid Films, manuscript accepted with major revisions

2 M. Villarica, M. J. Casey, J. Goodisman, J. Chaiken, J. Chem. Phys. 98 (1993) 4610-4625

3 S. Aceto, C. Y. Chang, R. W. Vook, Thin Solid Films, 219 (1992) 80-86

4 R. W. Vook, Materials Chemistry and Physics, 36(1994) 199-216

5 J. Goodisman, J. Chaiken, J. Photochem Photobio. A.

6 R. Botet and R. Jullien, J. Phys. Math. A: Math. Gen. 17 (1984) 2517-2530

7 John W. Brady, Jimmie D. Doll, Donald L. Thompson, J. Chem. Phys. 74(1981)1026-1028 and earlier papers referenced therein

8 I. Langmuir, J. Meteorl., 5 (1948) 175-192

9 R. Jullien, New J. Chem. 14 (1990) 239

10 R. C. Srivastava, J. Atmos. Sci. 45 (1988) 1091-1092

11 W.H. Stockmayer, J. Chem. Phys. 11 (1943) 45

12 G. Korvin in "Fractal Models in the Earth Sciences" (Elsevier, New York, 1992) page 193 in connection withdiscussion of equation (3.1.7)

13 H.E. Stanley, Form: An introduction to self-similarity and fractal behavior. In: H. E. Stanely and N. Ostrowsky(Editors), "On Growth and Form. Fractal and Non-Fractal Patterns in Physics." Martin Nijhoff Publ., Dordrecht-Boston-Lancaster, pp. 21-53

14 G. Korvin in "Fractal Models in the Earth Sciences" (Elsevier, New York, 1992) page 171

15 G. Korvin in "Fractal Models in the Earth Sciences" (Elsevier, New York, 1992) page 66

16 Yolanda J. Kime and Peter Grach, accepted Proc. Mat. Res. Soc.(Spring 1994 Meeting)

17 Robert W. Thomas and Donald W. Calbrese, PhenomenologicalObservations of Electromigration, 21st annual meeting proceedings, Reliability Physics, 1983, pp. 1-9

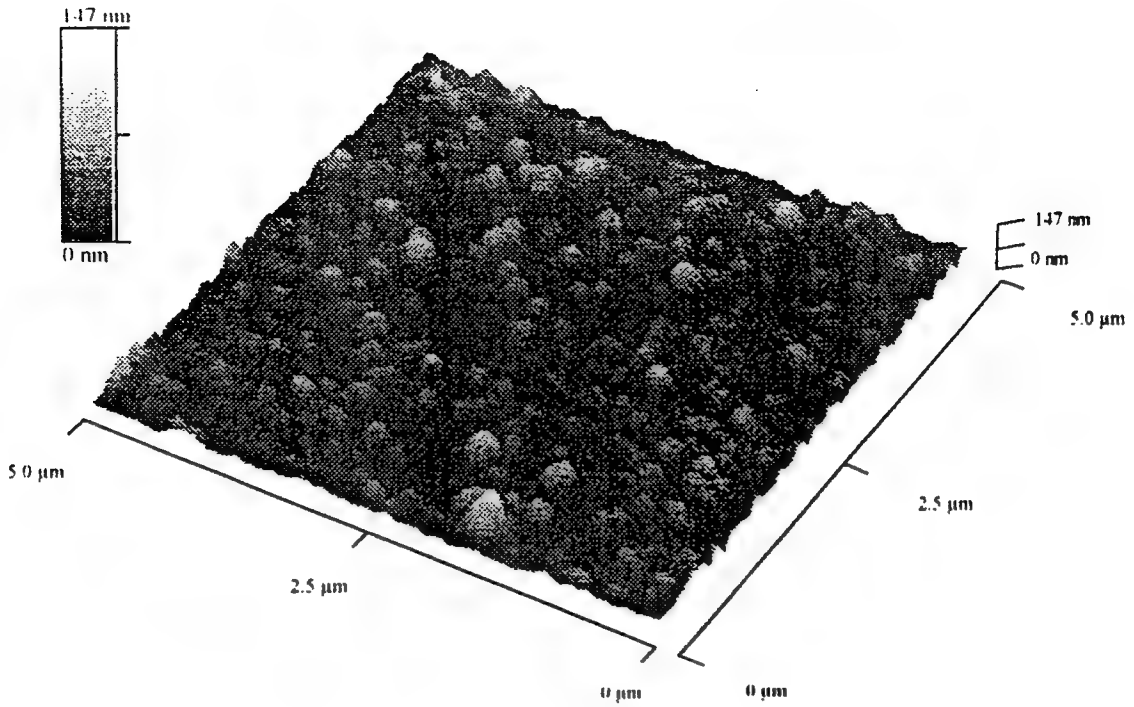


Figure 1a: 5µx5µ AFM image, lowest resolution (100x100 points).

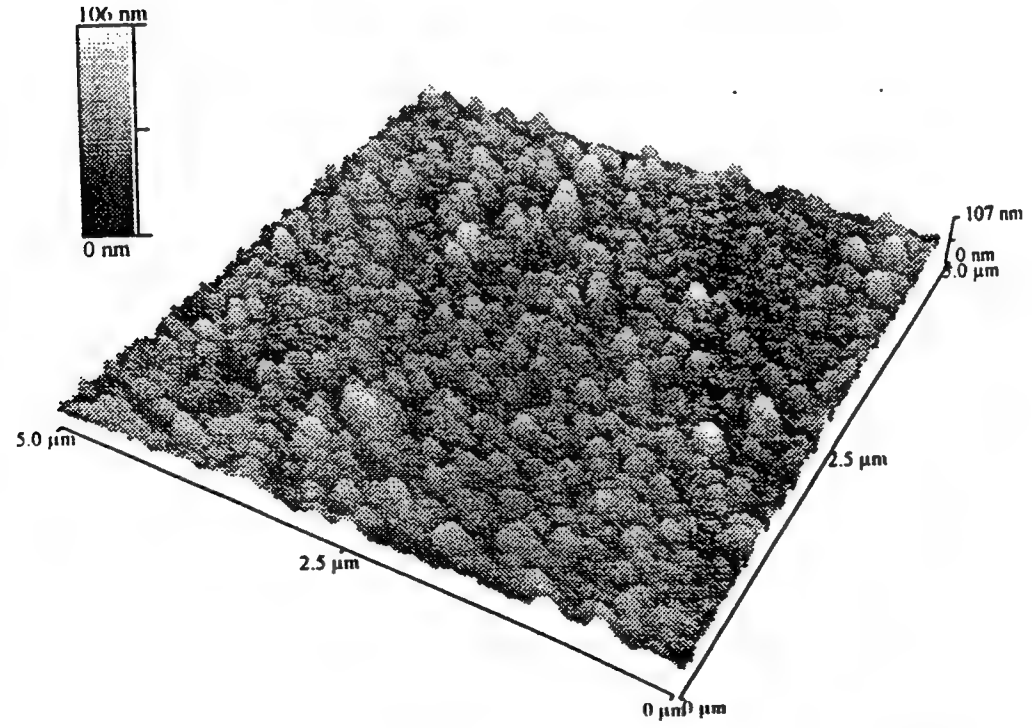


Figure 1b: 5µx5µ AFM image, highest resolution (500x500 points).

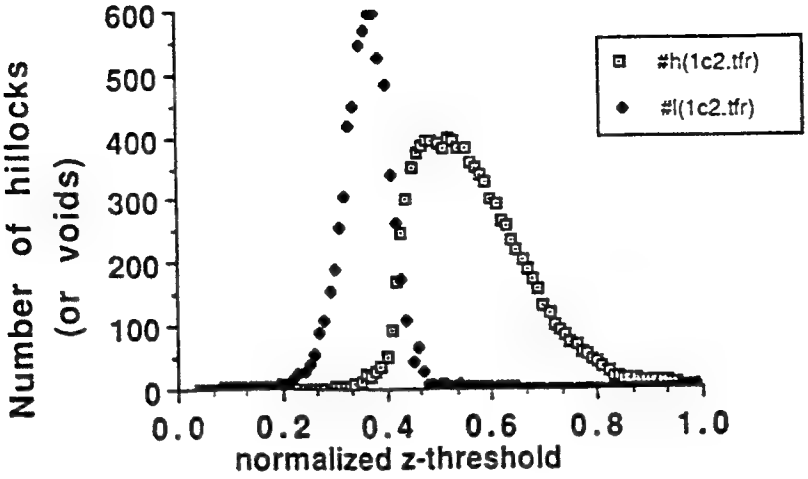


Figure 2: Number of hillocks (or voids) found as a function of z-threshold.

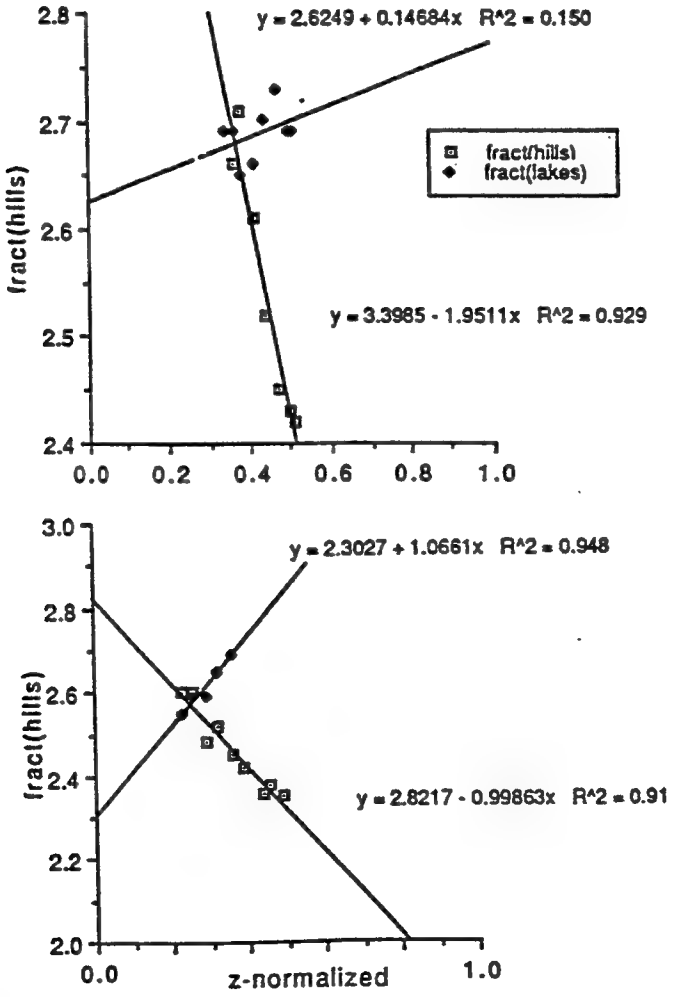


Figure 3a-b: Fractal dimension found as a function of z-threshold for 2 typical films. Some systematic variation seems apparent.

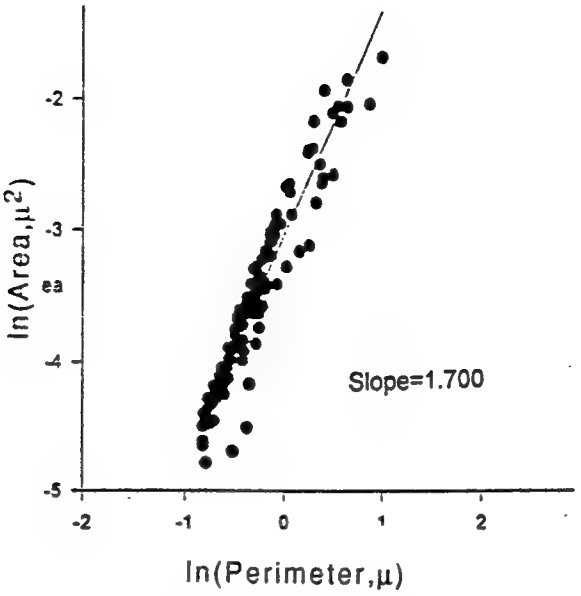


Figure 4: Typical area-perimeter plot of hillocks obtained from an AFM image.

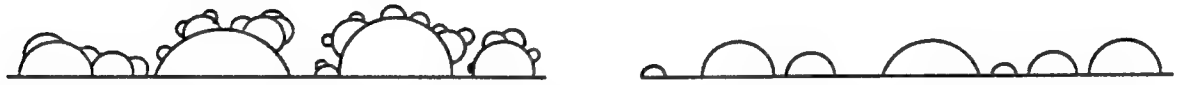


Figure 5: Surface profile for close packed hillocks (a) and non-overlapping hillocks (b).

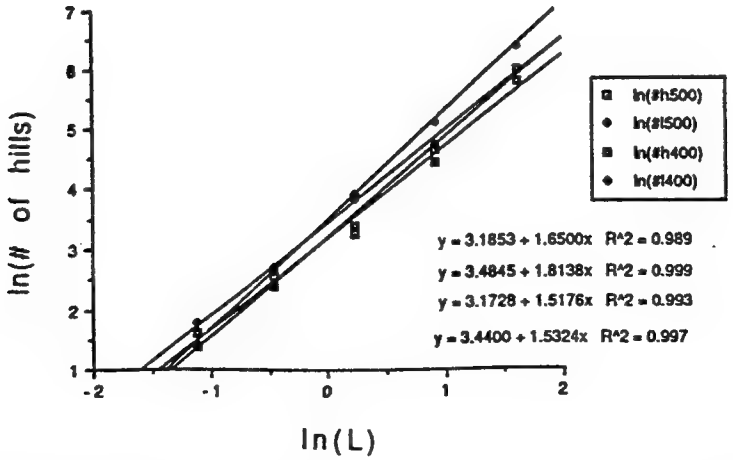


Figure 6: Number of hillocks (lakes) found for a particular z-threshold as a function of size of the region (L) examined. 500x500 and 400x400 point resolution data are used. Ordering of the fits matches the ordering of the legend.

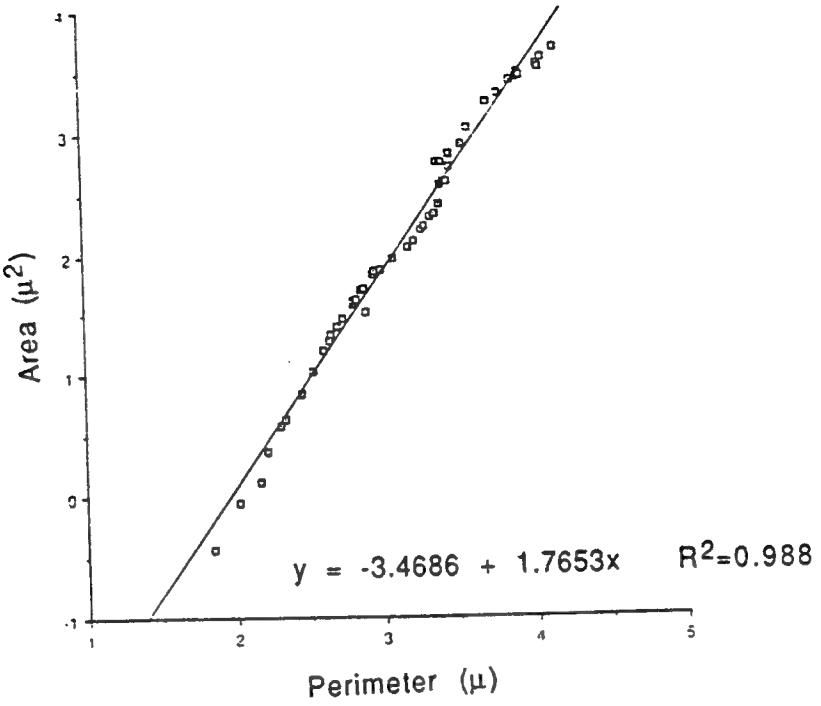


Figure 7: Area perimeter plot obtained from Professor Kime's data (see text).

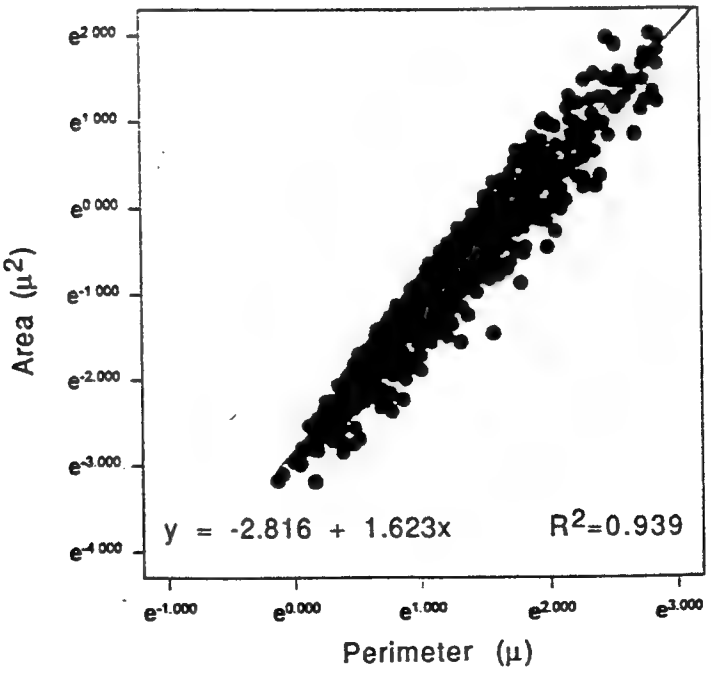


Figure 8: Area perimeter plot obtained from Thomas and Calabrese data (see text).

AN INVESTIGATION OF CEPSTRUM BASED SPEAKER IDENTIFICATION ALGORITHMS TO DETERMINE THEIR DEPENDENCY ON THE SPOKEN LANGUAGE

Stanley J. Wenndt
Graduate Student
Department of Electrical Engineering

Colorado State University
Electrical Engineering
Fort Collins. Colorado 80523

Final Report for:
Graduate Student Research Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

September 1994

AN INVESTIGATION OF CEPSTRUM BASED SPEAKER IDENTIFICATION ALGORITHMS TO DETERMINE THEIR DEPENDENCY ON THE SPOKEN LANGUAGE

Stanley J. Wenndt
Graduate Student
Department of Electrical Engineering
Colorado State University

Abstract

The objective of this research effort is to determine to what degree the Rome Lab cepstral based speaker identification algorithms are language dependent. The approach taken uses the same cepstral based techniques that Rome Laboratory/IRAA has successfully used for the automatic identification of speakers and applies them toward the automatic identification of languages. The speaker identification algorithms extract several cepstrum based features and feeds them to various classifiers. The results of the classifiers are then adjudicated to determine the final classification. Training and testing was completed for ten languages. Different speakers were used in training and testing to insure that the algorithms are targeting the languages and not speakers. Several observations provide insight into the relationship between the speaker identification techniques and their ability to automatically identify the languages tested.

I.) Introduction

This paper attempts to use several acoustic features (namely the cepstrum and its variations) for language identification. Applications, such as language translation, speech recognition, and speaker identification, could benefit greatly from a front end language identification system. For example, continuous speech recognition in an open environment (*i.e.*, the speaker may say anything in any language) is a near-impossible (if not impossible) task. However, if one can first identify the language, the speaker, and the environment the speaker operates in (*i.e.*, such as a F-16 cockpit), then one may have a realistic chance at identifying what is being said. The approach to this project was to use the same techniques that the Speech Processing Department at Rome Laboratory is using for speaker identification and apply them toward language identification using the OGI database. The goal was to determine to what degree the Rome Lab speaker identification algorithms are language dependent.

For speaker identification, each transmission for training and testing is divided up into frames of 256 samples where each frame is multiplied by a Hamming window. The frames are overlapped by 50% so that with a sampling rate of 8 KHz, the frames are 32 ms wide and advance in steps of 16 ms. The frames are used to calculate various features, namely the Cepstrum, Delta Cepstrum, Acceleration Cepstrum, Liftered Cepstrum, and RASTA-PLP Cepstrum. These features are then fed to several classifiers. The outputs of each classifier are fused together to try to exploit how different classifiers make different errors. While this is an simplistic view of speaker identification, the details and techniques will be described later. For a complete description of speaker identification, see [1], [2].

II.) Review of Past Work

Numerous efforts in language identification have been conducted over the years. What follows is a brief summary of some of the past work. Sugiyama [17] used acoustic features based on LPC

analysis. Two vector quantizer (VQ) classifiers were used. One where each language had its own codebook and one where all the languages together were assigned a codebook. The results yield recognition rates as high as 80%. This article was of particular interests because the techniques seemed to be quite similar. However, comparable results were not obtained with this research. Savic [16] used pitch contour analysis which measures the change in pitch vs. time along with Hidden Markov Models (HMM) to represent the state of the vocal tract. Foil [14] examined prosodic features extracted from pitch and energy contours along with formants. Foil had better success with the formant feature. K-means clustering algorithms and VQ classifiers were used. Goodman [15] improved on this work by using the first four formants and a Euclidean distance measure. Zissman [12] used HMM and Gaussian Mixtures. The features extracted included the mel-scale weighted cepstrum and the delta cepstrum. Zissman gives a more thorough review of previous works in this paper. Zissman [13] also uses phoneme recognition and language modeling for language identification.

III.) Acoustic Features

The main question that arises with parameter extraction is, "What is a cepstrum and why does it work?" A simplistic model of speech is $s(t) = e(t) * v(t)$ where $e(t)$ represents the vocal source or excitation which is rapidly varying, and $v(t)$ is the vocal tract which is slowing varying. Figure 1 depicts this simplistic speech production model. In the frequency domain, the spectrum of $s(t)$ becomes

$$S(\omega) = E(\omega)V(\omega) \tag{1}$$

By taking the logarithm of $S(\omega)$, the contributions due to the excitation and vocal tract become additive.

$$\log S(\omega) = \log E(\omega) + \log V(\omega). \tag{2}$$

When the inverse transform is taken of the logarithm spectrum, the vocal tract information tends to

occur near the origin while the excitation information shows up at multiples of the pitch period. The excitation of a periodic pulse trains is still a periodic pulse train even after the complex cepstrum, which is a nonlinear transformation, is taken. The mathematical proof of this follows.

Let

$$e(n) = \alpha_0\delta(n) + \alpha_1\delta(n - N_p) + \alpha_2\delta(n - 2N_p) + \alpha_3\delta(n - 3N_p) + ... \quad (3)$$

$$e(n) = \sum_{r=0}^{M} \alpha_r\delta(n - rN_p) \quad (4)$$

Taking the z-transform gives:

$$E(z) = \alpha_0 + \alpha_1 z^{-N_p} + \alpha_2 z^{-2N_p} + \alpha_3 z^{-3N_p} + ... \quad (5)$$

which can be written as

$$E(z) = a_0(1 - a_1 z^{-N_p})(1 - a_2 z^{-N_p})(1 - a_3 z^{-N_p})(1 - a_4 z^{-N_p}).... \quad (6)$$

$$E(z) = a_0 \prod_{r=1}^{M}(1 - a_r z^{-N_p}) \quad (7)$$

Taking the logarithm of $E(z)$ gives:

$$\hat{E}(z) = log(a_0) + \sum_{r=1}^{M} log(1 - a_r z^{-N_p}) \quad (8)$$

where

$$log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... \quad (9)$$

Therefore,

$$\hat{E}(z) = log(a_0) + \sum_{r=1}^{M} -a_r z^{-N_p} - \frac{a_r^2 z^{-2N_p}}{2} - \frac{a_r^3 z^{-3N_p}}{3} - .... \quad (10)$$

$$\hat{E}(z) = log(a_0) - \sum_{r=1}^{M}\sum_{l=1}^{\infty} \frac{a_r^l z^{-lN_p}}{l} \quad (11)$$

Taking the inverse z-transform of $\hat{E}(z)$ gives:

$$\hat{e}(n) = log(a_0)\delta(n) - \sum_{r=1}^{M}\sum_{l=1}^{\infty} \frac{a_r^l}{l}\delta(n - lN_p) \quad (12)$$

For $n = 160$ and $N_p = 80$,

$$\hat{e}(160) = -\sum_{r=1}^{M}\sum_{l=1}^{\infty}\frac{a_r^l}{l}\delta(160 - 80l) \tag{13}$$

The second summation is non-zero only for $l = \frac{160}{80} = 2$. Note that l is always an integer.

$$\hat{e}(160) = -\sum_{r=1}^{M}\frac{a_r^2}{2} \tag{14}$$

Generalizing gives:

$$\hat{e}(n) = log(a_0) \tag{15}$$

for $n = 0$ and

$$\hat{e}(n) = -\frac{1}{l}\sum_{r=1}^{M}a_r^l \tag{16}$$

for $n = N_p, 2N_p, 3N_p, ...$ $(l = \frac{n}{N_p} = 1, 2, 3...)$.

Equation 12 of $\hat{e}(n)$ represent the complex cepstrum of a periodic pulse train. It is obvious that $\hat{e}(n)$ is also a periodic pulse train with the same spacing.

Taking the inverse transform of the logarithm of the transform is officially termed the "cepstrum." When only magnitude information is present, it is termed the real cepstrum. When phase information is included, it is termed the complex cepstrum. For an excellent description, summary, and graphical explanation of the cepstrum, see [5]. Also, see [6], [7], [8], [9].

Working in the z-domain, the complex cepstrum is defined as the inverse z-transform of the logarithm of the z-transform of $s(n)$ where $s(n)$ is the sampled version of $s(t)$. We can solve for $\hat{c}(n)$ by using an all-pole model for speech, $S(z) = \frac{Q}{A(z)}$. The LPC coefficients, which model the spectrum, are calculated by using the autocorrelation method and the Levinson-Durbin equations. The natural logarithm is assumed.

$$\hat{C}(z) = \log S(z) = \log\frac{Q}{A(z)} = \log Q - \log A(z) \tag{17}$$

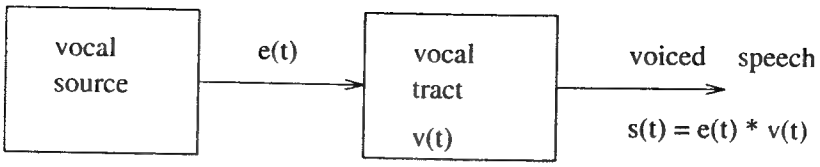


Figure 1: Simplistic speech production model

To solve for $\hat{c}(n)$, start by taking the derivative of both sides and multiply by $zA(z)$.

$$\frac{\partial \hat{C}(z)}{\partial z} = \frac{1}{A(z)} \frac{-\partial A(z)}{\partial z} \tag{18}$$

$$\frac{z\partial \hat{C}(z)}{\partial z} A(z) = \frac{-z\partial A(z)}{\partial z} \tag{19}$$

Take the inverse z-transform which gives:

$$-n\hat{c}(n) * a(n) = na(n) \tag{20}$$

Expanding the convolution and rearranging terms yields:

$$-\sum_{k=0}^{n-1}(n-k)\hat{c}(n-k)a(k) = na(n) \tag{21}$$

$$-n\hat{c}(n) - \sum_{k=1}^{n-1}(n-k)\hat{c}(n-k)a(k) = na(n) \tag{22}$$

$$\hat{c}(n) = -a(n) - \sum_{k=1}^{n-1}\frac{n-k}{n}\hat{c}(n-k)a(k) \tag{23}$$

For $n = 0$, $\hat{c}(0) = logQ$ and $n < 0$, $\hat{c}(n) = 0$. For $n > 0$, the recursive equation of $\hat{c}(n)$ is

$$\hat{c}(n) = -a(n) - \sum_{k=1}^{n-1}\frac{n-k}{n}\hat{c}(n-k)a(k) \tag{24}$$

Or, if the convolution is expanded differently, then

$$\hat{c}(n) = -a(n) - \sum_{k=1}^{n-1}\frac{k}{n}\hat{c}(k)a(n-k) \tag{25}$$

Two other extensions of the cepstrum coefficients are the Delta Cepstrum and the Acceleration Cepstrum. The Delta Cepstrum is the first derivative of the Cepstrum and the Acceleration Cepstrum is the second derivative of the Cepstrum. The derivatives are calculated by using a N-point linear regression over each cepstral channel. Both features provide information about spectral changes that have occurred between frames.

Liftering is used to try to separate the excitation and vocal tract information. Lowpass liftering is used to recover the vocal tract information while highpass liftering is used to recover the excitation information. Liftering can also be used to compute a weighted distance. A weighted Euclidean distance is used for these experiments [5].

The PLP Cepstrum and the RASTA-PLP Cepstrum attempt to compensate for the fact the LPC coefficients model all frequency equally well. This is contrary to human hearing since the human auditory system does not perceive pitch in a linear fashion. The major steps involved in calculating the PLP-Cepstrum are: 1.) Critical band analysis where the power spectrum is warped and down-sampled to the Bark scale. 2.) Equal loudness curves which compensates for the fact that the sensitivity of human hearing is not equal at all frequencies. 3.) Intensity power law which takes into account that the perceived loudness of a sound is different than its actual intensity. For details and explanations of the PLP-Cepstrum see [10]. The RASTA-PLP Cepstrum is an extension of the PLP Cepstrum. It attempts to compensate for channel induced distortions. See [11] for details. All cepstrum vectors are 14^{th} order.

IV.) OGI Database

The OGI database (Oregon Graduate Institute Multi-language Telephone Speech Corpus [18]) is a ten language database. According to [19], the National Institute of Standards and Technology has designated the OGI database as the standard for evaluating language identification algorithms. The ten languages are: English, Farsi, French, German, Japanese, Korean, Mandarin, Spanish,

Tamil, and Vietnamese. The data was gathered over the telephone by having native speakers of each language answer various questions. Speech was recorded at 8 KHz with 14 bit resolution. Since the data is "real-world", *i.e.*, not recorded in a controlled environment, many sentences contain pauses, breath noises, giggles, etc. Files ending in .adc contain the recorded speech. There is a header at the beginning of each file which contains the following information:

Variable Name	Variable Type	Description of Variable	# of Bytes
ad_hdrsize	short int	size of the header in short words	2
ad_version	short int	version of the header	2
ad_channels	short int	number of channels	2
ad_rate	short int	sample delta-t in $\frac{1}{4}\mu$secs	2
ad_samples	int	number of data samples (each sample is a short int)	4
little_indian	int	true if least significant byte is 0, *i.e.*, Vax byteorder	4

Table 1. Header information on *.adc files from the OGI database.

Before using a file, one should check the header to make sure the information is correct. If the header size is 0 or the sample delta-t is not equal to 500, it would be wise to discard the file. In case clarification on the delta-t notation is needed, here's how it works. Given a sample rate of 8 KHz, the sample period is, $\tau = \frac{1}{8000} = 125$ μsecs. Since there are four $\frac{1}{4}\mu$secs in a μsec, simply multiply τ by 4. This yields, $\tau = 500$ $\frac{1}{4}\mu$secs.

V.) Experiments

For the training data, 10 speakers of each language were used. The speaker was being asked to describe something about his/her home town that he/she liked (the files end with .htl.adc where the "htl" stands for "home town likes"). Only the first 200 vectors were used from each training file. This way, there was an equal number of training vectors for each language. This results in 6.4 seconds of speech for each speaker and 64 seconds of speech for each language. The training

files were taken from the */devtest directory. For testing, 10 speakers of each language were also used, but the data was taken from the */finaltest directory. Therefore, all the test speakers were different from the training speakers. This ensured speaker and text-independent results. The entire utterance was used for testing.

For both training and testing data, the data is divided up into frames of 256 samples with a 50 % overlap. This corresponds to a window size of 32 *m*secs which advances in steps of 16 *m*secs. Each frame is classified as speech/nonspeech with the nonspeech segments being discarded. The length of the test files varied from about 1 second of actual speech to 14 seconds. The speech frames are fed through a feature extractor which calculates each of the 5 cepstrum parameters. The features are then fed to a K nearest neighbor (KNN) classifier, a vector quantizer (VQ) [3], or a multilayer perceptron (MLP) [4].

VI.) Results of the KNN classifier

The KNN classifier compares each 14^{th} dimensional vector to its k nearest neighbors based on a Euclidean distance. The class most often associated with the test vector is selected. k was set to 9 for these experiments. Table 2 gives the results of the KNN classifier for the different features.

Language	Accel.	Cepst.	Delta	Lifter	Rasta
English	60	30	50	40	30
Farsi	10	0	10	10	30
French	0	70	20	70	30
German	10	0	0	0	0
Japanese	40	10	40	10	40
Korean	20	40	20	50	10
Mandarin	20	40	50	50	50
Spanish	0	10	0	10	10
Tamil	0	10	0	10	30
Vietnamese	0	10	0	10	10
Average	**16**	**22**	**19**	**26**	**24**

Table 2. Classification results using a KNN classifier with $k = 9$.

VII.) Results of the VQ classifier

The VQ classifier starts by computing 1 codeword for all the training vectors. Then it splits the codeword in two. The size of the codebook continues to double by splitting until a certain number of codewords are reached or when an error criteria is met. The number of codewords was set to 1024. This gave the best results compared to codebooks with fewer codewords. Due to the software, 1024 was the maximum number of allowable codewords. This is a large amount of codewords compared to the speaker identification algorithm which uses about 40 codewords. This large number of codewords seems to indicate tremendous confusibility in the data. In other words, these features may work well for speaker identification, but do not work well for language identification. Table 3 gives the results of the VQ classifier for the different features.

Language	Accel.	Cepst.	Delta	Lifter	Rasta
English	40	20	60	40	40
Farsi	0	10	30	0	0
French	10	80	0	50	40
German	20	20	30	0	0
Japanese	40	0	30	10	30
Korean	20	40	20	0	10
Mandarin	30	40	40	30	80
Spanish	10	10	0	10	20
Tamil	0	40	10	50	10
Vietnamese	0	10	10	20	30
Average	**17**	**27**	**23**	**21**	**26**

Table 3. Classification results using a VQ classifier with 1024 codewords.

VIII.) Results of the MLP classifier

The design of a multi-layer network is for interactive learning. For a given training input, the output is computed and compared to the expected output. Depending on the error, the weights are adjusted accordingly. This is repeated till a desired error rate is reached or until the network gets stuck in a local minima. For a 2-layer network (only 1 hidden layer), the number of hidden

nodes was first set to 30. However, the network would quickly get stuck in a local minima with an training error rate of about 70 %. Out of desperation, the number of hidden nodes was increased to 103. The training error rate dropped to about 55 %, but obviously, this is still very poor. Once again. this seems to demonstrate that these features do a poor job of capturing things unique to each language. Table 4 gives the results of the MLP classifier for the different features.

Language	Accel.	Cepst.	Delta	Lifter	Rasta
English	60	20	10	30	30
Farsi	0	0	0	0	20
French	0	50	0	60	50
German	0	30	0	50	0
Japanese	0	10	0	10	0
Korean	0	10	100	20	30
Mandarin	70	10	40	30	20
Spanish	10	40	10	20	10
Tamil	50	30	0	20	30
Vietnamese	30	10	0	10	60
Average	**22**	**21**	**16**	**25**	**25**

Table 4. Classification results using a MLP (14-103-10) classifier.

IX.) Adjudication of the Classifiers

In an attempt to exploit the different errors that the different classifiers make and/or the differents errors that result from the different features, some fusion techniques were applied. During fusion, each test file is normalized by the language with the highest number of winning frames. The normalized scores from each classifier are then added together, with the high score being the winner. Consider the data below in Table 5 from an English test file. Here, the VQ classifier was being used for the Liftered, Delta, and RSTA-PLP Cepstrum features. The score below each language represents the number of frames classified as that language.

Feature	Engl.	Farsi	Fren.	Germ.	Japa.	Kore.	Mand.	Span.	Tamil	Viet.
Liftered	5	4	4	0	7	4	2	8	11	7
Delta	13	5	6	2	3	9	2	2	5	5
RASTA	8	1	14	0	8	0	7	8	3	3

Table 5. Raw scores of a test file for 3 features with the VQ classifier.

The data is normalized as shown below in Table 6. Although the Delta Cepstrum was the only feature to identified the test file correctly, the overall score still selects English. The Liftered and RASTA feature helped contribute to this even though they didn't select English as the winner.

Feature	Engl.	Farsi	Fren.	Germ.	Japa.	Kore.	Mand.	Span.	Tamil	Viet.
Liftered	0.45	0.36	0.36	0.00	0.64	0.36	0.18	0.73	1.00	0.64
Delta	1.00	0.38	0.46	0.15	0.23	0.69	0.15	0.15	0.38	0.38
RASTA	0.57	0.07	1.00	0.00	0.57	0.00	0.50	0.57	0.21	0.21
Total Score	**2.02**	0.81	1.82	0.15	1.44	1.05	0.83	1.45	1.59	1.23

Table 6. Normalized scores and the overall score for the same test file.

The following tables summarize some adjudication results. For Table 7, each feature was fed to the 3 different classifiers and the results were adjudicated. For Table 8, the 5 features were fed to one type of classifier and the results were adjudicated. Many combinations of classifiers and features were tried, but few improved the results and some decreased the overall performance. The best results came from adjudicating the Liftered, Delta, and RASTA features which had all been fed through a VQ classifier. This yielded an identification rate of 31%. Obviously, when the identification rate is very poor to begin with, the benefits of adjudication will probably be minimal.

Language	Accel.	Cepst.	Delta	Lifter	Rasta
English	80	50	50	40	40
Farsi	0	0	10	0	20
French	0	60	0	60	50
German	0	10	0	10	0
Japanese	40	10	20	10	20
Korean	10	40	50	20	30
Mandarin	50	40	60	40	60
Spanish	0	20	0	20	20
Tamil	20	20	0	20	40
Vietnamese	20	10	0	10	20
Average	**22**	**26**	**19**	**23**	**30**

Table 7. Adjudication of the VQ, MLP, and KNN classifier for the 5 cepstrum features.

Language	VQ	MLP	KNN
English	70	50	70
Farsi	10	0	10
French	80	40	60
German	0	10	0
Japanese	10	20	20
Korean	20	20	10
Mandarin	60	50	60
Spanish	10	30	10
Tamil	30	40	10
Vietnamese	10	20	0
Average	**30**	**28**	**25**

Table 8. Adjudication of the 5 cepstrum features for the VQ, MLP, and KNN classifier.

X.) Further Analysis of the Results

The scores in Tables 2–4 only indicate if the test file was scored correctly or not. For further analysis, it is of interest to see how often the test file came in 2^{nd} or 3^{rd} place. In other words, if a English test file wasn't scored correctly, did the number of frames labeled as English come in 2^{nd} or 3^{rd} place? Tables 9–11 summarizes these results for the 3 classifiers using the RASTA feature. The RASTA feature was chosen because it had the highest average for the 3 classifiers.

Language	1^{st} place finishes	2^{nd} place finishes	3^{rd} place finishes	Total
English	3	3	0	6
Farsi	3	2	0	5
French	3	3	4	10
German	0	1	3	4
Japanese	4	0	3	7
Korean	1	3	1	5
Mandarin	5	1	0	6
Spanish	1	1	0	2
Tamil	3	0	0	3
Vietnamese	1	1	2	4

Table 9. 1^{st}, 2^{nd}, and 3^{rd} place finishes using the RASTA feature with the KNN classifier.

Language	1^{st} place finishes	2^{nd} place finishes	3^{rd} place finishes	Total
English	4	1	3	8
Farsi	0	2	3	5
French	4	3	2	9
German	0	3	0	3
Japanese	3	1	1	5
Korean	1	3	3	7
Mandarin	8	0	0	8
Spanish	2	1	1	4
Tamil	1	0	2	3
Vietnamese	3	2	2	7

Table 10. 1^{st}, 2^{nd}, and 3^{rd} place finishes using the RASTA feature with the VQ classifier.

Language	1^{st} place finishes	2^{nd} place finishes	3^{rd} place finishes	Total
English	3	3	2	8
Farsi	2	1	0	3
French	5	1	0	6
German	0	2	1	3
Japanese	0	3	2	5
Korean	3	0	3	6
Mandarin	2	3	0	5
Spanish	1	0	1	2
Tamil	3	1	2	6
Vietnamese	6	0	2	8

Table 11. 1^{st}, 2^{nd}, and 3^{rd} place finishes using the RASTA feature with the MLP classifier.

XI.) Conclusion

The goal of this effort is to determine to what degree the cepstral based speaker identification algorithms are language dependent. Examination of the scores produced by the five features tested shows that none of the features provided good language identification results. It is also interesting to note that no one feature is consistent across all the languages. The acceleration cepstra obtained the lowest score indicating it contained the least amount of information about the languages. The English and Mandarin languages were the most recognizable languages with these features while Farsi and German did consistently poor.

Results for the speaker identification algorithm were obtained by adding classifiers and adjudicating the results. This improved the language recognition results to a dismal average of 31%. Although the tests were limited in size, the results show a trend indicating that the cepstral based speaker identification algorithms are relatively independent of the spoken language. However more testing using multilingual speakers will be required to show the degree to which the speaker identification algorithms are dependent on the language spoken.

The fact that the overall language identification accuracy is poor makes intuitive sense. The cepstrum features provide information about the vocal excitation and vocal tract. This information is more unique to each speaker rather than the language spoken. Consequently, it makes sense that these features fail to accurately distinguish between languages.

Testing using multilingual speakers could have been performed to determine how much the speaker identification algorithms are effected by the spoken language. However this author is not aware of a data base that contains many speakers that speak many languages, and such a data base is not easily and inexpensively developed.

References

[1] Booz•Allen & Hamilton, Inc., "Speaker Identification Technology Program", Final Report Contract No. F30602-93-C-0011.

[2] R. Ricart, J. Cupples, L. Fenstermacher, "Speaker Recognition in Tactical Communications", *ICASSP '94 Proceedings*, vol. 1, pp. 329-332, April 1994.

[3] Y. Linde, A. Buzo, R. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Trans. Communication*, COM-28, pp. 84-95, January 1980.

[4] J. Hertz, et al. (1991) *Introduction To The Theory of Neural Computation*, Addison-Wesley Publishing Company, The Advanced Book Program, 350 Bridge Parkway, Redwood City, CA 94065.

[5] J. Deller, et al, (1993) *Discrete-Time Processing of Speech Signals*, New York: Macmillan Publishing Company, ch. 6, pp. 352-405.

[6] A. Oppenheim, R. Schafer, "Homomorphic Analysis of Speech", *IEEE Trans. Audio Electroacoustic*, vol. AU-16, pp. 221-226, June 1968.

[7] A. Oppenheim, "Speech Analysis-Synthesis System Base on Homomorphic Filtering", *J. Acoust. Soc. Am.*, vol. 45, no. 2, pp. 458-465, 1969.

[8] A. Noll, "Cepstrum Pitch Determination", *J. Acoust. Soc. Am.*, vol. 41, pp. 293-309, February 1967.

[9] B. Bogert, M. Healy, J. Tukey, "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking", *Proc Symp. on Time Series Analysis*, M. Rosneblatt, Editor, New-York: Wiley, ch. 15, pp. 109-243.

[10] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech", *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp. 1738-1752, April 1990.

[11] H. Hermansky, N. Morgan, A. Bayya, P. Kohn, "RASTA-PLP Speech Analysis Technique". *ICASSP '92 Proceedings*, vol. 1, pp. 121-124, March 1992.

[12] M. Zissman, "Automatic Language Identification Using Gaussian Mixture and Hidden Markov Models". *ICASSP '93 Proceedings*, vol. 2, pp. 399-402, 1993.

[13] M. Zissman, E. Singer "Automatic Language Identification of Telephone Speech Messages Using Phoneme Recognition and N-Gram Modeling", *ICASSP '94 Proceedings*, vol. 1, pp. 305-308, April 1994.

[14] J. Foil, "Language Identification Using Noisy Speech", *ICASSP '86 Proceedings*, vol. 2, pp. 861-864, April 1986.

[15] F. Goodman, A. Martin, R. Wohlford, "Improved automatic language Identification in Noisy Speech", *ICASSP '89 Proceedings*, vol. 1, pp. 528-531, May 1989.

[16] M. Savic, E. Acosta, S. Gupta, "An Automatic Language Identification System", *ICASSP '91 Proceedings*, vol. 2, pp. 817-820, May 1991.

[17] M. Sugiyama, "Automatic Language Recognition Using Acoustic Features", *ICASSP '91 Proceedings*, vol. 2, pp. 813-816, May 1991.

[18] Y. Muthusamy, R. Cole, B. Oshika, "The OGI Multi-language Telephone Speech Corpus". *Proc. Intl. Conf. Spoken Language*, October 1992.

[19] Y. Muthusamy, N. Jain, R. Cole, "Perceptual Benchmarks for Automatic Language Identification", *ICASSP '94 Proceedings*, vol. 1, pp. 333-336, April 1994.